DOI:
10.1038/s41588-018-0088-x

Document Version
Peer reviewed version

Regulatory variants at *KLF14* influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition

Kerrin S. Small[1*]†, Marijana Todorčević[2*], Mete Civelek[3,4*], Julia S. El-Sayed Moustafa[1*], Xiao Wang[5*], Michelle M. Simon[6], Juan Fernandez-Tajes[7], Anubha Mahajan[7], Momoko Horikoshi[2,7], Alison Hugill[8], Craig A. Glastonbury[1], Lydia Quaye[1], Matt J. Neville[2,9], Siddharth Sethi[6], Marianne Yon[8], Calvin Pan[4], Nam Che[4], Ana Viñuela[1], Pei-Chien C. Tsai[1], Abhishek Nag[1], Alfonso Buil[10], Gudmar Thorleifsson[11], Avanthi Raghavan[12], Qiurong Ding[13], Andrew Morris[7,14], Jordana T. Bell[1], Unnur Thorsteinsdottir[11,15], Kari Stefansson[11,15], Markku Laakso[16], Ingrid Dahlman[17], Peter Arner[17], Anna L. Gloyn[2,7,9], Kiran Musunuru[5*], Aldons J. Lusis[4,18,19*], Roger Cox[8*], Fredrik Karpe[2,9*], Mark I. McCarthy[2,7,9*]†

[1] Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
[2] Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK
[3] Center for Public Health Genomics, Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, USA
[4] Department of Medicine, University of California, Los Angeles, California, USA
[5] Cardiovascular Institute, Department of Medicine, Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA
[6] Biocomputing, MRC Harwell, Oxford, UK
[7] Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK
[8] Genetics of type 2 diabetes, Medical Research Council Harwell Institute, Oxford, UK
[9] Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK
[10] Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
[11] deCODE Genetics, Reykjavik, Iceland
[12] Harvard Medical School, Boston, Massachusetts, USA
[13] Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences ,Shanghai, ,PR China
[14] Department of Biostatistics, University of Liverpool, Liverpool, UK
[15] Faculty of Medicine, University of Iceland, Reykjavik Iceland
[16] Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
[17] Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden
[18] Department of Human Genetics, University of California, Los Angeles, California, USA
[19] Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California, USA

*these authors contributed equally to this work

†Corresponding authors
Correspondence should be addressed to K.S.S. (Kerrin.small@kcl.ac.uk) or M.I.M (mark.mccarthy@drl.ox.ac.uk).

Individual risk of type 2 diabetes (T2D) is modified by perturbations of adipose mass, distribution and function. To investigate mechanisms responsible, we explored the molecular, cellular, and whole-body effects of T2D-associated alleles near KLF14. We show that KLF14 diabetes-risk alleles act in adipose tissue to reduce KLF14 expression, and modulate, in trans, expression of >400 genes. We demonstrate that, in human cellular studies, reduced KLF14 expression increases pre-adipocyte proliferation but disrupts lipogenesis, and, in mice, adipose-specific deletion of Klf14 partially recapitulates the human phenotype of insulin resistance, dyslipidemia and T2D. We show that KLF14 T2D risk-allele carriers shift body fat from gynoid to abdominal stores, and display a marked increase in adipocyte cell size: these effects on fat distribution, and the T2D-association, are female-specific. Metabolic risk associated with variation at this imprinted locus depends on both the sex of the subject, and of the parent from whom the risk-allele derives.

The replicated genome-wide significant T2D association signal at chr7q32.3 maps to a 45kb recombination interval, extending from 3kb to 48kb upstream of KLF14[1,2] (Figure 1a-c). In previous work based on microarray-derived RNA expression data, KLF14, which encodes an imprinted transcription factor, was exposed as the likely cis-effector gene for this locus in subcutaneous adipose tissue[1] and revealed to be a trans-regulator of a programme of adipose tissue expression[3]. The KLF family of zinc-finger binding proteins have wide-ranging regulatory roles in biological processes such as proliferation, differentiation and growth[4,5]. However, little is known about KLF14, a single exon gene whose transcription is limited to the maternally inherited chromosome in embryonic, extra-embryonic, and adult tissue in humans and mice[6].

RESULTS

Adipose-specific regulation of KLF14 mediates the T2D association

Using RNAseq data from subcutaneous adipose biopsies collected from 776 female twin members of the TwinsUK cohort[7], we confirmed the cis-expression quantitative trait locus (eQTL) (using rs4731702 T2D risk allele C as the reference, β=-0.56, $P=1.8\times10^{-36}$) and identified a 38-fold-expanded trans-network of 385 genes (false discovery rate (FDR) <5%) (Supplementary Table 1, Figure 2). The KLF14 cis and trans effects were robust, replicating in three independent studies of subcutaneous adipose expression (MGH[8], METSIM[9] and deCODE[10]) (Figure 2c; Supplementary Table 1, Supplementary Table 2). Despite detectable KLF14 expression in a range of tissues, these cis- and trans-associations were completely adipose-specific with no eQTL signal evident in (a) skin, whole blood and lymphoblastoid cell lines (LCL) from the same TwinsUK individuals[7] (Figure 1f, Figure 2d); (b) T2D-relevant tissues such as muscle[11], liver[8,12], and islet[13]; or (c) the broader coverage represented in eQTL datasets such as GTEx[14] (Supplementary Table 3). KLF14 is imprinted, and as with the T2D-association[1], the KLF14 cis and trans eQTLs were maternal-specific – the paternally inherited allele had no effect on KLF14 expression (Supplementary Figure 1, Supplementary Table 4). All other genes within 300kb of the interval are also imprinted, but paternally-expressed, confirming KLF14 as the likely mediator of the T2D association at this locus.

We first sought to refine the location of the causal variant(s) responsible for these associations. The T2D association could not be resolved beyond a set of 29 SNPs in high mutual linkage disequilibrium (LD, r^2>0.94 in UK10K); the apparent European-specificity of the T2D association[15] precludes trans-ethnic fine-mapping. The ancestral T2D-risk allele varies in global frequency (EUR = 54%, ASN=69%, AMR=58%, AFR=73%) but we could detect no evidence of positive selection (Supplementary Figure 2, Supplemental Note). We utilized chromatin state maps to annotate the associated interval, identifying a ~1.6kb enhancer ~5kb upstream of KLF14. This enhancer encompasses five of the 29 associated variants and shows marked tissue-specificity: in ChromHMM predictions from ENCODE[16] and Roadmap[17], it is annotated as active in 10 (out of 127) cell types and tissues, three of them derived from adipose tissue, including adipocytes and adipose-derived mesenchymal stem cells (Figure 1d). The enhancer is also active in certain blood cells, but there was no corresponding cis-eQTL[18-20], making it unlikely that transcriptional regulation in blood cells contributes to the GWAS phenotypes. There were no other adipose active enhancers in the region.

As methylation is one of the key processes related to enhancer function, we explored population-level methylation data at this locus in many of the same TwinsUK individuals. The T2D-risk haplotype was associated with increased methylation levels at Illumina 450K array probe cg02385110, ~3kb upstream of KLF14, in subcutaneous adipose tissue (n=603, $P=2.2x10^{-7}$, β=0.01 for rs4731702) but not whole blood (n=309, P=0.69) or skin (n=437, P=0.39) (Figure 1e-i). There was a consistent direction of effect across these analyses in adipose: the T2D risk-haplotype was associated with increased methylation and decreased RNA expression, and cg02385110 methylation and KLF14 expression were negatively correlated. A second probe (cg08097417), located at the KLF14 transcription start site has been associated with age in whole blood[21] and adipose tissue samples[22]. However, this probe lies outside the T2D-association interval, and cg08097417 methylation was not related to KLF14 expression (P=0.36) or risk-haplotype (P=0.99), indicating that age-related variability of cg08097417 is unrelated to KLF14 expression or disease pathogenesis (Supplementary Figure 3). Instead, we conclude that T2D-associated risk attributable to this locus is likely to be consequence of sequence variation at the adipose enhancer upstream of KLF14 and is marked by altered methylation.

The KLF14 variants regulate a large adipose-specific trans network

The trans-network regulated by the KLF14 variants is remarkable both for its size and robust replication. Consistent with the known function of KLF14 as both transcriptional activator and repressor, the trans-associations included both positive and negative effects (Figure 2a, Supplementary Table 1). Mediation analysis applied to the RNA expression data supported a causal role for KLF14 expression in regulation of many of the trans-genes (161/385 genes passed Bonferroni-corrected Sobel's mediation $P<1.3x10^{-4}$).

The principal trans-regulatory mechanism appears to be direct interaction of KLF14 with trans-gene cis-regulatory elements; the 20kb regions upstream of

the 385 trans genes were enriched both for KLF14 binding peaks in empirical CHiPSeq data[23] from HEK293T cells ($P=2x10^{-4}$) and the presence of the proposed KLF14 binding motif (Normalized Enrichment Score=4.17, $P=1.5x10^{-5}$). We assessed functional annotation of the trans-genes using ToppGene[24], and found the subset of 177 trans-genes with KLF14 binding-motifs to be enriched for "metabolic pathways" ($q=1.2\times10^{-3}$) and "binding by PPARG and RXRA during adipocyte differentiation" ($q=3.9x10^{-8}$) (Supplementary Table 5). Further subsetting, defining 122 trans-genes enriched for both KLF14 binding motifs and KLF14 ChIP-seq peaks revealed broadly similar functional enrichment ("binding by PPARG and RXRA during adipocyte differentiation": q=0.04; "metabolic pathways": $q=2.6x10^{-4}$).

A sub-network of 18 trans-genes showed enrichment for SREBF1 binding motifs (NES=4.22, $P=1.2x10^{-5}$), 11 of which did not have KLF14 motifs (Figure 2b). The sub-network of SREBF1 motif-containing trans-genes was enriched for "cholesterol biosynthesis and lipid metabolism" ($q=2.8\times10^{-7}$) (Supplementary Table 5). SREBF1, a transcription factor involved in cholesterol homeostasis, is itself a trans-gene directly regulated by KLF14 and appears to act as an intermediary in the regulation of this sub-network.

The human trans-network includes several genes with functions that make them attractive candidates for mediating the range of KLF14-associated phenotypes. These include SLC2A4 and IDE, which encode the GLUT4 transporter and insulin degrading enzyme, respectively. GLUT4 mediates glucose uptake in adipose tissue and skeletal muscle, among other tissues, and adipose-specific reduction in SLC2A4 expression levels has been reported in T2D[25]. IDE degrades peptides including insulin, glucagon, and amylin, and maps to a long-established T2D genome-wide association study (GWAS) interval[15]. Notably, at multiple trans-genes, we identified adipose cis-eQTLs coincident with metabolic-trait GWAS SNPs ($r^2>0.8$) including three known T2D loci (STARD10, C6orf57 and CDK2AP1) (Supplementary Table 6). This shows that local regulation of these genes can mediate T2D-susceptibility independent of trans-regulation via KLF14, and indicates that the phenotypic consequences of KLF14 variation are likely mediated by multiple genes within the trans network.

Sex-specific metabolic trait associations implicate insulin action

Previous studies have demonstrated that T2D-risk alleles at KLF14 are, in non-diabetic individuals, associated with increased fasting insulin[26] and reduced high density lipoprotein (HDL)-cholesterol[27,28], emphasizing a primary impact on insulin action rather than insulin secretion. By collecting data from the largest-available GWA meta-analyses for multiple traits[26,28-31], we have extended the spectrum of KLF14 association to encompass a broad range of insulin resistant, "metabolic syndrome" phenotypes including low density lipoprotein (LDL)-cholesterol, triglycerides, waist-hip ratio and the HOMA-IR measure of insulin sensitivity[32] (Table 1). Crucially, as demonstrated in Shungin et al[31], the impact of KLF14 variation was far more marked on fat distribution rather than overall adiposity (as measured by body mass index (BMI)), with a particularly strong association between the T2D-risk allele and reduced hip circumference ($\beta=-0.017$, $P=1.6x10^{-6}$).

Sex-stratified reanalysis of these GWAS data revealed that, for all traits with available data, effect sizes in females substantially exceeded those in males (Table 1). The associations with hip circumference, triglycerides, LDL and fasting insulin were exclusive to females, while the HDL, WHR and T2D associations displayed a strong female-bias (e.g. T2D: female $P=2.2x10^{-6}$, odds ratio (OR) 1.14 [1.08-1.20]; male P=0.002, OR 1.08 [1.03-1.14]). In UKBiobank (N=118,192), we found significant SNPxSex interactions for Hip Circumference ($P=3.1x10^{-7}$) and Waist Hip Ratio adjusted for BMI ($P=3.9 \times 10^{-6}$), but not for Waist Circumference (P=0.06) or BMI (P=0.43). In a combination of UKBiobank, Wellcome Trust Case Control consortium and The Resource for Genetic Epidemiology on Adult Health and Aging (GERA) samples (13,728 cases; 129,911 controls), rs4731702 and sex interacted to influence T2D-risk ($P=6.6x10^{-3}$).

These analyses demonstrate that the metabolic and anthropometric consequences of KLF14 variation are dependent on both the parental origin of the risk-allele and the sex of the recipient of that allele. Based on combining these sex-specific ORs with the known imprinting-related maternal-specificity of these associations[1], we estimate that the point estimate for the per-allele OR for T2D is ~1.28 in women inheriting the risk-allele from their mother. In contrast, although KLF14 expression in adipose tissue was higher in females than males across all genotype classes, we found no equivalent sex-difference in the magnitude of the adipose cis-eQTL effect (Figure 3a, Supplementary Table 4). To further explore drivers of the sexual dimorphism in KLF14 expression levels, we compared KLF14 expression level between 86 pre-menopausal and 487 post-menopausal females from TwinsUK. Individuals taking hormone replacement therapy were excluded from this analysis. There was no association between KLF14 expression and menopausal status (P=0.81) and the cis-eQTL effect size in each group was comparable, suggesting gonadal steroids do not underlie the observed sexual dimorphism in KLF14 expression levels. We conclude that the sex-dependency of the whole body phenotype is not the direct consequence of sexual dimorphism at the cis-regulatory level.

The genetic associations were consistent with our observations using gene expression levels in the TwinsUK cohort. KLF14 gene expression in adipose tissue was associated with a combined insulin resistance phenotype characterized by increased fasting insulin and triglycerides, and reduced HDL cholesterol ($P=1.08x10^{-3}$). KLF14 trans-genes also showed enrichment for association between trans-gene expression levels and the same combined insulin resistance phenotype ($P=1.82x10^{6}$) (Supplementary Table 7- 8). We used Dual energy X-ray Absorptiometry (DXA)-derived body composition data from the same TwinsUK individuals to further dissect the impact on fat distribution. Whilst there was no association between KLF14 expression in subcutaneous adipose tissue and total fat volume (P=0.56) or BMI (P=0.27), there were differences with respect to the distribution of that fat, including an inverse association with abdominal visceral fat (P=0.02) and with the ratio of android:gynoid fat ($P=1.7x10^{-3}$). In many other insulin-resistant settings, abnormalities of fat distribution are associated with ectopic deposition of fat

including the liver, leading to non-alcoholic fatty liver disease. However, we found no evidence that KLF14 risk variants influence liver fat deposition, whether measured directly (P=0.26, N=7,176)[33] or indirectly via disturbed liver function (alanine aminotransferase (ALT) P=0.79, gamma-glutamyl transferase (GGT) P=0.89, N=61,089)[34].

These data indicate that genetically-determined reduction in cis-expression of KLF14 in adipose results in an insulin-resistant and T2D-predisposing phenotype, which, in women only, is characterized by, and potentially mediated by, a shift in fat distribution from relatively inert gynoid stores to more metabolically-active abdominal deposition.

Conserved mouse phenotypes

We generated mice carrying an adipose-specific Klf14 deletion allele through the combination of a CRISPR-Cas9-generated conditional allele with Klf14-flanking loxP sites and adipose-specific (Adipoq) Cre recombinase and compared these mice with wildtype Cre-expressing colony mates. Adipose-specific knockout mice of both sexes (null on both maternal and paternal chromosomes) displayed a broad pattern of insulin resistance phenotypes: HDL-C was reduced (16 weeks, significant only in females), triglycerides increased (16 weeks, significant only in males) (Figure 4a,b), and both glucose tolerance and insulin sensitivity were impaired in both sexes at 12 weeks (Figure 4c-f).

We also characterised two mouse-lines in which Klf14 expression was reduced globally. The first mouse ($Klf14^{tm1(KOMP)Vlcg}$; generated within KOMP) featured deletion of the entire single exon of Klf14. The second mouse, designed to minimise any disruption of adjacent regulatory sequences, used CRISPR-Cas9 genome editing to induce a 7bp frameshift indel allele. There were some inconsistencies between the models, but the predominant phenotype was a reduction in HDL-C (Supplementary Figures 4a and 5a) with a modest and/or transitory effect on glucose homeostasis (Supplementary Figures 4b,c,d and 5d,e). The HDL-C phenotype seen in global deletion mice mirrors the phenotype reported following hepatic-specific deletion of Klf14[35].

RNASeq analysis of subcutaneous fat taken from both the adipose Klf14 KO and the $Klf14^{tm1(KOMP)Vlcg}$ global deletion mouse recovered some, but not all, features of the trans-network seen in humans. In the adipose-specific Klf14 KO mouse (N=8), 18 genes were differentially expressed at FDR $q<0.05$ (or 1286 with $p<0.05$). The latter set (of 1286 nominally significant mouse trans-genes) was significantly enriched for overlap with the human trans-network ($P<2x10^{-4}$). In the $Klf14^{tm1(KOMP)Vlcg}$ global deletion mouse (N=16), we confirmed this overlap ($P<1x10^{-6}$ with the human trans network) in the 285 genes differentially expressed compared to controls (FDR $q<0.05$): 5' noncoding sequences for these genes were also enriched for the presence of the Klf14 binding motif ($P<0.001$). Although Srebf2 (the mouse ortholog of SREBF1) is directly regulated by Klf14, we did not detect enrichment of the murine trans-genes for the Srebf2 motif. The murine trans-genes were enriched for relevant functional categories, including "response to lipid" ($q=4.4x10^{-5}$), "triglyceride

metabolic process" ($q=2.3x10^{-3}$), "response to steroid hormone" ($q=2.7x10^{-3}$) and "regulation of cell proliferation" ($q=3.2x10^{-3}$) (Supplementary Table 9).

Thus, adipose-specific deletion of Klf14 in mice recapitulates many aspects of the phenotype associated with variants that influence adipose-specific cis-regulation of KLF14 in humans, but the absence of sex specificity, and differences in the molecular consequences of Klf14 perturbation place limits on the relevance of murine models for this gene.

KLF14 impacts glucose uptake, lipogenesis and cell size

Members of the KLF family are involved in transcriptional control of adipocyte development and function[5], but the role of KLF14 in adipogenesis is unknown. To evaluate the role of KLF14 in adipocyte development, we measured KLF14 expression in primary pre-adipocytes isolated from human abdominal subcutaneous adipose tissue biopsies (female N=4, male N=4) during proliferation and subsequent differentiation over 14 days. Adipocyte expression of KLF14 was higher in females compared to males at all time-points (Figure 3b), demonstrating that the sex-differential expression of KLF14 observed in adult biopsies is present throughout adipogenesis.

We next investigated the link between reduced KLF14 expression and adipocyte development. Compared to tissue from non-risk allele homozygotes, fresh adipose tissue explants from female risk allele homozygotes showed a 44% reduction (P=0.001, N=132) in lipogenesis as measured by the incorporation of ^{14}C- glucose label into the triacylglycerol backbone (Figure 5e). There was no significant difference in males (P=0.49, N=32) (Figure 5e). We confirmed these findings in vitro through shRNA knockdown (40% at day 1 and 33% at day 14) of KLF14 (Figure 5a) in primary pre-adipocytes isolated from abdominal subcutaneous biopsies (N=7 females): this resulted in 60% reduction in TAG accumulation after 14 days of differentiation (P=0.03) (Figure 5c). The adipocyte differentiation wholly relies on de novo lipogenesis as the cell culture medium contains glucose but not fatty acids. Knowing that SLC2A4 (which encodes GLUT4) is a KLF14 transgene, we first tested the expression of GLUT4 in the KLF14 shRNA differentiated adipocytes. This showed prominent reduction (by 64±7% (P=0.02)) (Figure 5a). To investigate the functional consequences of the reduction in GLUT4 expression, we quantified the insulin-stimulated glucose uptake, finding a 50% reduction in KLF14 shRNA differentiated adipocytes (Figure 5d). Concomitant with the lower glucose uptake and lipogenesis, the expression of adipocyte maturation factors PLIN1 (-42%±8, P=0.02) LEP (-54%±13, P=0.03), and FITM2 (-12%±4, P=0.04) was significantly ablated in the knockdown cells, whilst the levels of the classical early-differentiation, pro-adipogenic transcription factors (CEBPA, PPARG2) were not changed (Figure 5a). KLF14 knockdown also increased cellular proliferation, observed as reduced doubling time (P=0.02) (Figure 5b). We infer that reduced levels of KLF14 result in a defect in glucose uptake resulting in impaired adipocyte lipogenesis and maturation, and propose that the increase in proliferation results from fruitless cycling at earlier stages of differentiation.

Adipocyte size is an important marker of adipose tissue dysfunction and metabolic disease and, in some studies, has been predictive of insulin resistance and T2D independent of obesity[36-40]. We assessed adipocyte size and cell number distribution by histological analysis of subcutaneous abdominal and gluteal adipose tissue biopsies from sex-, age- and BMI-matched pairs of individuals from the Oxford Biobank, homozygous for either the risk or non-risk haplotypes (N=18 males, 18 females). At both sites, there was a marked shift towards increased adipocyte size in T2D-risk haplotype homozygotes: this was only observed in women (comparison of medians: abdominal P=0.008, gluteal P=0.02) (Figure 6b,d,e; Supplementary Figure 6a,b,c). We estimated, assuming spherical cell morphology, that there was a two-fold difference in adipocyte volume between the genotypic groups in females. Gluteal and abdominal tissue from female T2D-risk haplotype homozygotes also contained fewer total adipocytes (comparison of medians: Abdominal TT=$2.21\pm0.31x10^{10}$, GG=$0.66\pm0.19x10^{10}$, P=0.04; Gluteal TT=$4.87\pm0.93x10^{10}$, GG=$2.25\pm0.58x10^{10}$, P=0.04). The cell size findings and their sex-specificity were confirmed using a different technique (measuring diameter of isolated adipocytes from collagenase digested tissue) in an independent cohort of 28 females (P=0.004), and 16 males (P=0.78) also matched for age and BMI (Figure 6c).

DISCUSSION

T2D risk variants that lead to reduced expression of KLF14 in adult adipose tissue are therefore also implicated in a defect of adipogenesis that is likely to reflect impaired glucose uptake. The consequence of this defect appears to be a profound effect on subcutaneous adipose tissue structure characterised by the presence of fewer but larger adipocytes. We infer that KLF14 risk-allele carriers are obliged to favour expansion of existing adipocytes to meet lipid storage needs, resulting in suboptimal fat storage and an increased risk of insulin resistance and T2D. This hypothesis of impaired fat storage is consistent with the elevated circulating triglyceride levels observed in KLF14 risk haplotype carriers. Reduced expression of KLF14 in female carriers of the risk-allele is likely to drive a shift in the distribution of adipose tissue between subcutaneous fat depots, favouring abdominal over gluteal deposition. This results in loss of the relative protection against the metabolic consequences of adipose tissue accumulation associated with adipose deposition in gynoid sites which has been widely observed in epidemiological studies[41,42] and is consistent with the observed association of the KLF14 T2D-risk allele and human fat distribution.

We note that the sex-specificity of the KLF14 risk alleles varies across biological domains: effects on adipocyte development and whole body phenotypes are more marked in females despite the fact that cis and trans regulatory effects are shared between the sexes. The simplest explanation for this discordance is a threshold effect driven by the higher baseline expression levels of KLF14 in females, though it is also possible that males lack downstream processes (e.g. related to hormonal environment) that mediate the response to KLF14. While the KLF14 T2D association was discovered through typical GWAS approaches, the remarkable specificity of this association – with respect to sex, parent of origin, ethnicity and tissue of

action - is a reminder that risk prediction, or targeted medical treatment, based on genotype alone may fail to capture highly-relevant aspects of biological complexity.

Acknowledgements
This study was supported by MRC grant MR/J010642/1 to K.S.S., R.D.C., F.K. and M.I.M.. K.S.S. is supported by a MRC New Investigator Award (L01999X/1). M.I.M. is a Wellcome Senior Investigator and is supported by Wellcome (090532, 106130, 098381, 203141), NIDDK (U01-DK105535) and the MRC (MR/L020149/1). M.C. is supported by NIH Award R00 HL121172. A.L.G. is a Wellcome Senior Fellow in Basic Biomedical Science (095101/Z/10/Z). K.M. is supported by NIH award R01-DK099571 and A.J.L. by NIH award NIH P01HL28481. R.D.C. is supported by MRC MC_U142661184. J.F. was a Marie Curie Fellow. A.P.M. is a Wellcome Senior Fellow is Basic Biomedical Science (award WT098017). M.L. is supported by Academy of Finland grants 77299 and 124243; the Finnish Heart Foundation; the Finnish Diabetes Foundation; Finnish Funding Agency for Technology and Innovation (TEKES) contract 1510/31/06; Commission of the European Community HEALTH-F2-2007-201681. A.V. and A.B. were supported by the European Union Framework Programme 7 grant EuroBATS (259749). Some computations were performed at the Vital-IT center for high-performance computing of the Swiss Institute of Bioinformatics (SIB; http://www.vital-it.ch/). The TwinsUK study was funded by Wellcome and European Community's Seventh Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. We acknowledge excellent technical support for animal husbandry (Mary Lyon Centre), genotyping, histology and pathology. We thank the volunteers from the Oxford Biobank, NIHR Oxford Biomedical Research Centre, for their participation. The Oxford Biobank (www.oxfordbiobank.org.uk) is also part of the NIHR National Bioresource which supported the recalling process of the volunteers. GERA data came from a grant, the Resource for Genetic Epidemiology Research in Adult Health and Aging (RC2 AG033067; Schaefer and Risch, PIs) awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical endonte

Author contributions
K.S.S., A.L.G., K.M., A.J.L., R.C., F.K., and M.I.M. conceived and designed the project M.T., M.C., J.S.E.M., M.M.S., J.F., C.A.G., L.Q., C.P., G.T., P.C.T. and A.N. analysed data. M.T., X.W, Q.D., A.R., A.H., S.S., M.Y. and N.C. performed experiments. A.M., M.H., M.J.N. and A.V. contributed experimental and technical support as well as discussion. A.B., J.T.B., I.D., P.A., M.L., U.T. and K.S. contributed data. K.S.S., M.T., M.C., J.S.E.M., M.M.S., K.M., A.J.L., R.C., F.K., and M.I.M. wrote the manuscript. All authors read and approved the manuscript.

Competing financial interests statment
M.I.M. has received consulting and advisory board honoraria from Pfizer, Lilly, and NovoNordisk. G.T., U.T. and K.S. are employees of deCODE Genetics/Amgen.

References

1 Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868-874, doi:nature08625 [pii]
10.1038/nature08625 (2009).

2 Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* **42**, 579-589, doi:ng.609 [pii]
10.1038/ng.609.

3 Small, K. S. *et al.* Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* **43**, 561-564, doi:ng.833 [pii]
10.1038/ng.833.

4 Dang, D. T., Pevsner, J. & Yang, V. W. The biology of the mammalian Kruppel-like family of transcription factors. *Int J Biochem Cell Biol* **32**, 1103-1121, doi:S1357-2725(00)00059-5 [pii] (2000).

5 Wu, Z. & Wang, S. Role of kruppel-like transcription factors in adipogenesis. *Developmental biology* **373**, 235-243, doi:10.1016/j.ydbio.2012.10.031 (2013).

6 Parker-Katiraee, L. *et al.* Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLoS Genet* **3**, e65, doi:06-PLGE-RA-0297R3 [pii]
10.1371/journal.pgen.0030065 (2007).

7 Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* **47**, 88-91, doi:10.1038/ng.3162 (2015).

8 Greenawalt, D. M. *et al.* A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res* **21**, 1008-1016, doi:10.1101/gr.112821.110 (2011).

9 Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *American journal of human genetics* **100**, 428-443, doi:10.1016/j.ajhg.2017.01.027 (2017).

10 Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423-428, doi:nature06758 [pii]
10.1038/nature06758 (2008).

11 Keildson, S. *et al.* Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. *Diabetes* **63**, 1154-1165, doi:10.2337/db13-1301 (2014).

12 Innocenti, F. *et al.* Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet* **7**, e1002078, doi:10.1371/journal.pgen.1002078 (2011).

13 van de Bunt, M. *et al.* Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. *PLoS Genet* **11**, e1005694, doi:10.1371/journal.pgen.1005694 (2015).

14 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

15 Replication, D. I. G. et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet **46**, 234-244, doi:10.1038/ng.2897 (2014).
16 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. Nature **489**, 57-74, doi:10.1038/nature11247 (2012).
17 Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. Nature **518**, 317-330, doi:10.1038/nature14248 (2015).
18 Fairfax, B. P. et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet **44**, 502-510, doi:10.1038/ng.2205 (2012).
19 Naranbhai, V. et al. Genomic modulators of gene expression in human neutrophils. Nature communications **6**, 7545, doi:10.1038/ncomms8545 (2015).
20 Rotival, M. et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet **7**, e1002367, doi:10.1371/journal.pgen.1002367 (2011).
21 Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Molecular cell **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).
22 Ronn, T. et al. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Human molecular genetics **24**, 3792-3813, doi:10.1093/hmg/ddv124 (2015).
23 Najafabadi, H. S. et al. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nature biotechnology **33**, 555-562, doi:10.1038/nbt.3128 (2015).
24 Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res **37**, W305-311, doi:10.1093/nar/gkp427 (2009).
25 Garvey, W. T. et al. Pretranslational suppression of a glucose transporter protein causes insulin resistance in adipocytes from patients with non-insulin-dependent diabetes mellitus and obesity. The Journal of clinical investigation **87**, 1072-1081, doi:10.1172/JCI115068 (1991).
26 Horikoshi, M. et al. Discovery and Fine-Mapping of Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation. PLoS Genet **11**, e1005230, doi:10.1371/journal.pgen.1005230 (2015).
27 Global Lipids Genetics, C. et al. Discovery and refinement of loci associated with lipid levels. Nat Genet **45**, 1274-1283, doi:10.1038/ng.2797 (2013).
28 Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature **466**, 707-713, doi:nature09270 [pii]
10.1038/nature09270.
29 Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. Nature **518**, 197-206, doi:10.1038/nature14177 (2015).
30 Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet **44**, 981-990, doi:10.1038/ng.2383 (2012).

31 Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature **518**, 187-196, doi:10.1038/nature14132 (2015).

32 Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet **42**, 105-116, doi:ng.520 [pii]
10.1038/ng.520.

33 Speliotes, E. K. et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet **7**, e1001324, doi:10.1371/journal.pgen.1001324 (2011).

34 Chambers, J. C. et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet **43**, 1131-1138, doi:10.1038/ng.970 (2011).

35 Guo, Y. et al. Perhexiline activates KLF14 and reduces atherosclerosis by modulating ApoA-I production. The Journal of clinical investigation **125**, 3819-3830, doi:10.1172/JCI79048 (2015).

36 Arner, E. et al. Adipocyte turnover: relevance to human adipose tissue morphology. Diabetes **59**, 105-109, doi:10.2337/db09-0942 (2010).

37 Hammarstedt, A., Graham, T. E. & Kahn, B. B. Adipose tissue dysregulation and reduced insulin sensitivity in non-obese individuals with enlarged abdominal adipose cells. Diabetology & metabolic syndrome **4**, 42, doi:10.1186/1758-5996-4-42 (2012).

38 Lonn, M., Mehlig, K., Bengtsson, C. & Lissner, L. Adipocyte size predicts incidence of type 2 diabetes in women. FASEB journal : official publication of the Federation of American Societies for Experimental Biology **24**, 326-331, doi:10.1096/fj.09-133058 (2010).

39 Lundgren, M. et al. Fat cell enlargement is an independent marker of insulin resistance and 'hyperleptinaemia'. Diabetologia **50**, 625-633, doi:10.1007/s00125-006-0572-1 (2007).

40 Weyer, C., Foley, J. E., Bogardus, C., Tataranni, P. A. & Pratley, R. E. Enlarged subcutaneous abdominal adipocyte size, but not obesity itself, predicts type II diabetes independent of insulin resistance. Diabetologia **43**, 1498-1506, doi:10.1007/s001250051560 (2000).

41 Snijder, M. B. et al. Independent and opposite associations of waist and hip circumferences with diabetes, hypertension and dyslipidemia: the AusDiab Study. International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity **28**, 402-409, doi:10.1038/sj.ijo.0802567 (2004).

42 Yusuf, S. et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. Lancet **366**, 1640-1649, doi:10.1016/S0140-6736(05)67663-5 (2005).

43 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. Nature methods **9**, 215-216, doi:10.1038/nmeth.1906 (2012).

Figure Legends

Figure 1 |Cell-type-specific enhancer in the risk haplotype regulates *KLF14* expression. Association of KLF14 locus variants with (a) type 2 diabetes (N=69,033) (Morris et al[39]), (b) high density lipoprotein levels (N=99,900) (Teslovich et al[27]); (c) KLF14 mRNA abundance in adipose tissue of 776 women (TwinsUK). Circles represent genotyped and imputed DNA variants and are colored by LD r^2 values with the index SNP rs4731702. The red asterisk indicates the position of methylation probe cg02385110. Association between genotype and KLF14 expression was assessed using mixed effects models, correcting for family structure and relevant covariates (methods). Uncorrected P-values displayed. (d) Chromatin state annotations for the locus across 93 reference epigenomes (rows) for cell and tissue types profiled by the Roadmap Epigenomics and ENCODE Projects. Colors correspond to chromatin states; yellow=enhancer, red=promoter as detailed in Ernst et al[43]. ESC, embryonic stem cell; HSC, hematopoietic stem cell; iPSC, induced pluripotent stem cell. Blue box contains the putative enhancer. Chromosome coordinates correspond to UCSC Genome Browser build hg19. (e-i) Boxplots of KLF14 mRNA abundance (e-f) and methylation of probe cg02385110 (g-i) in concurrently-sampled adipose ($N_{Expression}$=720; $N_{Methylation}$=595), skin ($N_{Expression}$=606; $N_{Methylation}$=414) and whole blood ($N_{Expression}$=368; $N_{Methylation}$=289) from the TwinsUK study participants, based on rs4731702 genotype (CC=risk-allele homozygotes). KLF14 expression was undetectable in whole blood. Boxplots display the median, with hinges corresponding to the first and third quartiles. Whiskers extend from the hinge to the largest value no further than 1.5xinter-quartile range, with values beyond shown as outliers. Associations assessed using linear regression, correcting for family structure and relevant covariates (methods).

Figure 2 | *KLF14 cis*-eQTL is an adipose-specific *trans*-regulator of a large network of genes. a, Genomic location of the 385 trans-genes. Line colour indicates direction of trans-effect; blue, positive association with T2D risk-allele; green, negative. **b,** 177 trans-genes are enriched for KLF14 motifs and 18 trans-genes form a sub-network regulated by SREBF1. Genes are colored as in a) to indicate direction of trans-effect. Solid lines link KLF14 to trans genes containing KLF14 binding sites, dashed lines link SREBF1 to trans-genes containing SREBF1 binding sites **c,** KLF14 trans-network replicates in three independent adipose cohorts, METSIM (N=770), MGH (N=701) and deCODE (N=589). Histograms show p-value distribution of trans-genes in each replication cohort. **d,** Trans-network is not present in other tissues in the same TwinsUK samples as adipose discovery. Histograms show P-value distribution of trans-genes in skin (N=716), LCLs (N=814), and whole blood (N=384). In TwinsUK, METSIM and DeCODE, association between KLF14 lead SNP rs4731702 genotype and gene expression was assessed using linear models with correction for relevant covariates (see methods), and using the Kruskal-Wallis test in the MGH dataset. Uncorrected P-values displayed. Trans-gene associations were considered replicated at $P<0.05$.

Figure 3 | *KLF14* expression is sex-differentiated in biopsies and throughout adipocyte differentiation. a, Adipose expression of KLF14 in the deCODE cohort. The KLF14 cis-eQTL has a similar effect size in males (N=265) and females (N=376), but KLF14 expression is higher in females than males across all genotype classes. Plots display KLF14 mean relative expression ±SEM. **b,** Pre-adipocytes derived from abdominal biopsies and subjected to a 14 day-differentiation protocol show higher KLF14 expression in females (N=4, mean±SEM) compared to males (N=4, mean±SEM) at all time points.

Figure 4 | Adipose-specific knockout of *Klf14* in the mouse. Clinical chemistry parameters were measured in female and male adipose-specific (Adipoq-Cre) knockout (KO) C57BL/6J Klf14 mice and their Adipoq-Cre-expressing, wildtype (Wt) controls. Mice were fed a standard diet throughout their lifetimes. **a**, HDL-C at 16 weeks significantly reduced in female KO group mice compared to wildtype (P= 0.04); **b**, TG at 16 weeks significantly increased in the male KO group compared to wildtype (P=0.03 unpaired two-tailed t-test); **c and d**, glucose levels in an IPGTT at 12 weeks of age significantly increased in female (c) and male (d) KO compared to wildtype controls at multiple time-points; **e and f**, in an ITT at 12 weeks of age glucose reduced to a lesser extent in KO mice compared to their wildtype controls at multiple time-points in both females (e) and males (f). Pairwise significance as compared to controls is shown. Analysis in (a) and (b) use unpaired two-tailed t-test (for male HDL-C, with Welch's correction), and in (c) to (f) by 2-way ANOVA with repeated measures and Bonferroni correction (GraphPad Prism 6). Wildtype mice shown as blue lines and fill, females (N=8) and males (N=8). Adipose KO mice shown in red lines and fill, females (N=5) and males (N=6). All error bars are plotted as a mean central value plus or minus the standard deviation.

Figure 5| *KLF14* expression effects expression of adipocyte maturation marker genes and adipocyte function. a, The mean fold change in expression of marker genes following KLF14 shRNA knockdown. Mean fold change is plotted as the expression of the labeled gene in KLF14 shRNA adipocytes divided by expression in control adipocytes. Expression of all genes was measured with Real Time PCR in seven paired knockdown/control replicates (N=7 ±SEM). A significant decrease in expression was seen for KLF14, measured at days 1 and 14. There was no change in early differentiation markers PPARG2 and CEBPA at either day 1 or 14. A significant decrease in expression was seen for late-stage differentiation markers LEP and PLIN1 and late-stage adipogenic trans-genes FITM2 and GLUT4 at day 14. Statistical significance was assessed using a Wilcoxon signed-rank two-sided test. (*) marks statistical differences between knockdown and control adipocytes for each gene ($p \leq 0.05$). **b,** KLF14 shRNA pre-adipocytes isolated from female abdominal adipose tissue biopsies (N=7, mean±SEM) have a significant decrease in doubling time compared to non-target shRNA controls. **c,** KLF14 shRNA differentiated adipocytes have a significant decrease in TAG accumulation (N=7, mean±SEM) and **d,** in insulin stimulated glucose uptake (N=6, mean±SEM) compared to non-target shRNA controls. Statistical significance was assessed using a Wilcoxon signed-rank two-sided test. **e,** Female KLF14 risk allele homozygotes (CC) have a significant decrease in lipogenesis, as measured in adipose explants, compared to non-risk allele

homozygotes (TT) (Female (N=132), Male (N=32), mean±SEM). The minimum and maximum values of lipogenesis are presented with their whiskers. The line in the box represents the median. Statistical significance was assessed using unpaired two-sided t-test.

Figure 6 | Adipose tissue of Type 2 diabetes risk allele homozygotes contains fewer, larger mature adipocytes compared to non-risk allele homozygotes. a, Sections of subcutaneous abdominal adipose tissue biopsies from two age- and BMI-matched female volunteers. Histological sections were stained with hematoxylin and eosin. **b,** The minimum and maximum values of adipocyte cell surface area in abdominal adipose histological sections stratified by sex and genotype are presented with their whiskers. The line in the box represents the median. >100 cells were measured for each biopsy (N=18 per sex). Statistical significance was assessed using a Wilcoxon signed-rank two-sided test. **c,** The minimum and maximum values ofadipocyte cell volume measured in dispersed adipocytes from collagenase-dispersed abdominal adipose tissue stratified by sex and genotype are presented with their whiskers (female N=28, male N=16, mean ± SEM). The line in the box represents the median. Statistical significance was assessed using a Wilcoxon signed-rank two-sided test. **d,** Cumulative frequency distribution of adipocyte cell surface area in females (N=18), measured as in **b**. **e,** Cumulative frequency distribution of adipocyte cell surface area in males (N=18), measured as in **b**.

Tables

Trait		Type 2 diabetes	Fasting insulin	Fasting glucose	HDL	TG	LDL	BMI	WHR	Waist	Hip
GWAS Citation		Morris, 2012	Horikoshi, 2015	Horikoshi, 2015	Teslovich, 2010	Teslovich, 2010	Teslovich, 2010	Locke, 2015	Shungin 2015	Shungin, 2015	Shungin, 2015
Sex-combined	β	1.10 (1.06-1.15)	0.019	0.007	-0.015	0.016	0.009	0.009	0.006	-0.009	-0.017
	P	$3.6x10^{-8}$	$1.2 x10^{-5}$	0.07	$1.2x10^{-15}$	$1.1 x10^{-6}$	0.02	0.002	0.08	0.009	$1.6 x10^{-6}$
	N	69,033	24,243	46,656	99,900	96,598	95,454	322,022	212,127	230,394	211,022
Female	β	1.14 (1.08-1.20)	0.031	0.011	-0.042	0.036	0.018	0.010	0.021	-0.008	-0.033
	P	$1.4x10^{-6}$	$2.4 x10^{-7}$	0.07	$3.5 x10^{-11}$	$1.0 x10^{-8}$	0.002	0.01	$3.6 x10^{-6}$	0.08	$9.9 x10^{-14}$
	N	40,413	13,073	23,663	62,816	59,473	61,803	171,899	117,967	126,971	117,288
Male	β	1.08 (1.03-1.14)	0.006	0.004	-0.034	0.013	0.010	0.009	0.011	-0.009	0.003
	P	0.003	0.36	0.58	$9.9 x10^{-7}$	0.04	0.19	0.03	0.02	0.04	0.50
	N	28,620	11,158	17,731	37,745	35,288	36,840	152,830	94,344	103,616	93,919

Table 1 | rs4731702 is associated with insulin resistance, lipid and body-composition phenotypes HDL, high density lipoprotein; TG, Triglycerides; LDL, low density lipoprotein; BMI, body mass index; WHR, waist-hip ratio. Morris et al[30] results are taken from the Stage 1 Discovery as Stage 2 included non-European samples. HDL and LDL effect sizes are reported in mmol/l, TG was log-transformed. Type 2 Diabetes effect sizes are Odds Ratios, not Betas.

Online Methods

Choice of lead SNP

The T2D association at the KLF14 locus consists of 29 SNPs in near perfect linkage disequilibrium (r^2>0.94). The lead SNP in T2D GWAS has been reported as rs4731702 (Kong, et al[1]) rs972283 (Voight, et al[2]) and rs10954284 (Morris et al[30] and Mahajan, et al[15]), all three of which are in perfect linkage disequilibrium in Europeans (r^2=1). The lead SNP for the HDL GWAS is rs4731702 (Teslovich et al[28], Willer et al[27]). As rs4731702 is present on all commonly used genotyping arrays we chose to report all results in this manuscript with respect to rs4731702, where C is the T2D risk allele and T is the non-risk allele.

TwinsUK gene expression

Biopsies and blood samples from 856 healthy female twins from the TwinsUK cohort were collected within the MuTHER project[44] and RNA sequenced as previously described[7]. See Supplementary Note for further details of the TwinsUK data.

Cis and trans eQTL analysis

Cis and trans eQTL analysis was conducted in the TwinsUK RNAseq data as follows. Exon read counts were corrected for technical covariates and family structure using a mixed effects model including insert size and mean GC content as fixed effects and primer index, sample processing date, family and zygosity as random effects. Corrected residuals were used for all eQTL analyses. Cis- and trans-eQTL associations were conducted using the MatrixeQTL package[45] using a standard additive linear model, with BMI, age and age-squared included as covariates. A window size of 1Mb was used for cis-eQTL analyses. A FDR[46] threshold of 5% was applied to trans-eQTL results, with FDR calculated using the qvalue package[47] in R version 3.1.1. The deCODE, METSIM and MGH cohorts were analyzed as previously described[8-10].

Functional fine mapping

We used the chromatin states predicted by ChromHMM[43] utilizing 127 reference epigenomes from the ENCODE[16] and Roadmap Epigenomics projects[55]. ChromHMM segments the genome into 15 states using 5 chromatin marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) with predictions ranging from active transcription start sites to repressed Polycomb complexes. The ChromHMM predicted enhancer spans a 1.6 kb region, ~4kb upstream of KLF14 transcription start site (TSS) and harbors 5 of the T2D GWAS SNPs (rs12154627, rs6973807, rs6974400, rs6974288, rs11762784).

TwinsUK methylation data

Adipose tissue sample in 603 individuals (taken from the same biopsy as the RNAseq data) and 309 whole blood samples from the TwinsUK study were profiled on the Illumina Infinium HumanMethylation450 BeadChip previously[48,49]. The following covariates were included in analyses, batch (beadchip), BS conversion efficiency (as assessed using the built-in BS conversion efficiency controls) and BS-treated DNA input. The association of

age with methylation was examined using a linear mixed effect regression model (LME) adjusting for batch effects, BMI, and, family and zygosity structure. Blood methylation was additionally corrected for estimated cell counts. To compare methylation and expression levels, each was first adjusted for covariates, and Pearson's correlation performed on the residuals.

Trans-network mediation analysis:
Significant mediation was determined by computing Sobel's test statistic[50]. To calculate the mediation score, the three following models were implemented:

Model 1:

Model 2:

Model 3:

Where Y = trans-gene expression, A = age, E = cis-gene expression, P = BMI, G = cis-genotype.

By conditioning on cis-gene expression (the mediator:) we can determine if each individual association detected in trans is regulated in cis or is independent by quantifying .

Enrichment of KLF14 Chipseq binding in trans-genes
We utilized the HEK293 KLF14 Chipseq data from Najafabadi et al (2015)[51]. Genomic co-ordinates of the peaks of KLF14 binding (along with the summit position for each peak) identified in the Chip-seq experiments were available in the supplementary data of Najafabadi et al (2015) (http://hugheslab.ccbr.utoronto.ca/supplementary-data/C2H2_B1H/). For the enrichment analysis for KLF14 binding, we utilised the summit position of the 18,652 distinct Chip-seq peaks of KLF14 binding (filtered on a minimum Phred-like quality score of call confidence of 50) provided in this data. An enrichment of KLF14 binding sites for the 385 genes with trans-expression association with the KLF14 locus (FDR < 5%) was evaluated by comparing the proportion of genes for which KLF14 binding sites were identified in the following two groups of genes: first, the 385 genes with trans-expression association with the KLF14 locus; and second, the remaining genes that were tested in the trans-expression analysis (N=18,678). An enrichment of KLF14 binding sites for the 385 genes with trans-expression association was further evaluated by performing 10,000 permutations of a random set of 385 genes sampled from the entire set of genes that was tested in the trans-expression analysis, and estimating the number of genes with KLF14 binding sites for each permutation.

iRegulon transcription factor binding analyses

In order to find the KLF14 regulon (a transcription factor (TF) and its direct transcriptional targets, which contain common TF binding sites in their cis-regulatory control elements) we use the iRegulon cytoscape plugin[52]. The prediction of regulons consists of four steps: 1) motif detection, this process relies on an offline scoring step of a sequence search space (10kb or 20kb around the TSS) whereby every gene in the human genome, along with orthologous sequences in ten other vertebrate genomes, is scanned with Cluster-Buster[53] for homotypic clusters of motifs using a library of nearly 10,000 position weight matrices (PWMs), resulting in a gene-ranking list for each PWM. Enriched motifs are those motifs for which the input genes are enriched at the top of the ranking, using the Area Under the Curve (AUC) of the cumulative curve; 2) Track discovery, this step also relies on an offline scoring step whereby every gene in the human genome is scored with around 1,120 ChiP-seq tracks, generating a list of TFs according to the highest ChiP peak within the regulatory space; 3) Motif2TF mapping, the candidate TFs are prioritized by finding the optimal path from a motif to a TF, in a motif-TF network; 4) Target detection, for each enriched motif, the candidate targets are selected as the significantly highly ranked genes compared to the genomic background and to the entire motif collection as background.

Functional enrichment analysis

KLF14 trans-genes were subdivided into three subsets based on the presence of bindings motifs for KLF14, SREBF1 or neither. The three subsets were assessed for functional enrichment using ToppFun, part of the ToppGene package[24]. Results were corrected for multiple testing using the Bonferroni method, with enrichments considered significant at Bonferroni-corrected $P < 0.05$.

Overlap between KLF14 trans-regulated genes and published GWAS results

The NHGRI-EBI GWAS catalog[54] (Available at: www.ebi.ac.uk/gwas, accessed November 2015) was filtered so as to retain signals meeting genome-wide significance ($P<5.0x10^{-8}$). Corresponding cis-eQTLs for each of the 385 KLF14 trans-genes, including secondary, tertiary and quaternary cis-eQTLs sequentially conditioned on the lead cis-eQTLs at each locus (a total of 465 cis-eQTLs), were then cross-referenced against the NHGRI GWAS results, retaining at each KLF14 trans-gene locus those reported NHGRI associations within +/-250kb of the relevant cis-eQTLs that showed moderate to high LD ($r^2 \geq 0.8$) between the reported GWAS SNP and the respective cis-eQTL.

Association between KLF14 expression and TwinsUK metabolic traits

To determine if gene expression was associated to a concurrently measured phenotype of interest, expression of each exon was treated as a quantitative trait in a linear mixed effects model implemented using the lme4 package[51]. The linear mixed effects model was adjusted for age, BMI and technical covariates (mean GC content and insert size mode) (fixed effects), family relationship (twin pairing), primer index and zygosity (random effects). Phenotypic data were treated as continuous independent traits, and were inverse normal transformed. A full model with the phenotype fitted was

compared to a null model (no phenotype) using a 1d.f. ANOVA. A FDR was estimated using the package QVALUE to obtain q-values that correspond to a controlled FDR 5%. TwinsUK metabolic phenotypes were measured at the same time-point as the biopsy and were collected as previously described[3]. Body-fat distribution traits were measured using dual-energy X-ray absorptiometry (DXA; Hologic QDR 4500 plus) with the standard protocol.

Trait-Expression Association Enrichment Analyses

The most significantly-associated exon per gene was retained for all genes. Enrichment was assessed by comparing the number of genes significantly associated with each phenotype (at $q<0.05$) within the set of KLF14 trans-genes ($N_{trans\text{-}genes}=385$) to the remaining genes not within the KLF14 trans-network ($N_{genes}=18,716$), using a Fisher's exact test.

Sex x SNP interaction analysis in large cohorts

Details of the data preparation, quality control and cohort specific analysis covariates for the UK Biobank, GERA and WTCCC/T2D data are provided in the Supplementary Methods. For evaluating the SNP x Sex interaction effect of rs4731702 on anthropometric traits in the UK Biobank, we tested for association of the genotype with each inverse normalized variable using linear regression with a main effect for SNP and SNP-by-sex interaction terms using an additive model. We adjusted for covariates: age, age^2, sex, six (within UK) ancestry principal components, and array used to measure genotypes. A maximum of 118,193 individuals (62,165 females and 56,027 males) with genotype and valid BMI, height, waist and hip were available for analysis. For evaluating the SNP-by-sex interaction effect of rs4731702 on risk of T2D, we considered a total of 13,728 T2D cases and 129,911 controls from 3 studies from populations of European ancestry. Sample and variant quality control was performed within each study (see Supplementary Methods). Within each study, the variant was tested for T2D association under an additive model after adjustment for study-specific covariates, including principal components to adjust for population structure and with a main effect for SNP and SNP-by-sex interaction terms. We then combined association summary statistics for the variant across studies via fixed-effects inverse-variance weighted meta-analysis.

Animal experiments information

CRISPR-Cas9 mice were generated in the Harvard University Genome Modification Facility and were kept and studied in accordance with Harvard University's Faculty of Arts and Sciences Institutional Animal Care and Use Committee (IACUC) protocol 14-05-202. Klf14$^{tm1(KOMP)Vlcg}$ mice were kept and studied in accordance with UK Home Office legislation and local ethical guidelines issued by the Medical Research Council (Responsibility in the Use of Animals for Medical Research, July 1993; home office license 30/3146).

Experimental design of in vivo mouse experiments in CRISPR-Cas9 mouse models

The CRISPR-Cas9 knockout mouse models were generated as described in the Supplementary Note. All procedures used for CRISPR-Cas9 animal studies were approved by Harvard University's Faculty of Arts and Sciences

Institutional Animal Care and Use Committee and were consistent with local, state, and federal regulations as applicable. All animals were housed in a 12-hour dark-light cycle and fed standard chow ad libitum in conditions similar to those described above. The control mice and KO mice were littermates from several litters that were timed to be the same age; the mice were kept in the same cages, in as few cages as permitted by the restrictions set forth by institutional policy (typically no more than three adult mice per cage).

For lipid measurements, blood samples were collected from the lateral tail vein following an overnight fast. Samples were kept on ice then centrifuged for 10 minutes at 2,000xg in a centrifuge at 4°C. The plasma levels of triglyceride, total cholesterol, and HDL-C were measured using Infinity Triglycerides Reagent (Thermo Fisher), Infinity Cholesterol Reagent (Thermo Fisher) and HDL-Cholesterol E (Wako Diagnostics) independently according to the manufacturers' instructions. For the intraperitoneal glucose tolerance test (IPGTT), the mice were fasted overnight, and glucose concentrations were measured immediately before and at 20, 40, 60, and 120 minutes after intraperitoneal injection of glucose (1g/kg body weight). For the insulin tolerance test (ITT), the mice were fasted for 4 hours, and glucose concentrations were measured immediately before and at 20, 40, 60, and 120 minutes after intraperitoneal injection with human insulin (Sigma) at 1U/kg body weight. Whole blood glucose levels were measured using a glucose meter and test strips (OneTouch). For histological examination, fat tissue was fixed in 4% paraformaldehyde and embedded in paraffin. Consecutive sections were cut and stained by hematoxylin and eosin for light microscopy examination and evaluation.

Experimental design of in vivo mouse experiments in deletion mouse models

$Klf14^{tm1(KOMP)Vlcg}$ deletion mice were generated as described in the Supplemental Note. Deletion mice were kept under controlled light (light 7am–7pm, dark 7pm–7am), temperature (21±2°C) and humidity (55±10%) conditions. They had free access to water (9–13 ppm chlorine) and were fed ad libitum on a commercial diet (SDS Rat and Mouse No. 3 Breeding diet, RM3) until 18 weeks of age when they were then transferred to a high fat diet (45% kcal from fat; D12451; Research Diets).

Cohorts of male and female mice were bred for longitudinal blood and body composition based phenotyping tests. These included four groups: 1) MAT (heterozygotes (hets) inheriting the allele maternally), 2) PAT (hets inheriting the allele paternally), 3) WT MAT controls and 4) WT PAT controls. Sample size estimates were based on previous experience and data from other mouse models in which the relevant traits were measured. These data were used in power calculations to help in deciding cohort sizes. Mice were housed in single sex groups of mixed genotype across multiple litters and were not randomised into groups. Mouse IDs and genotypes were recorded on the cages and were not blinded to the operator carrying out the animal procedure although subsequent tests only include animal ID information and clinical

chemistry was carried out by a core service without knowledge of the genotypes.

Echo-MRI analysis and fasted blood sampling from the lateral tail vein (local anaesthetic) were carried out at 8, 12, 22 and 27 weeks of age. Whole blood glucose was measured using an AlphaTRAK meter and test strips (Abbott). Plasma insulin was assayed using a Mouse insulin ELISA kit (Mercodia). Terminal blood samples were collected from 33 week old mice by intraperitoneal anaesthesia and cardiac puncture following an overnight fast. Samples were kept on ice then centrifuged for 10 minutes at 8,000xg in a centrifuge set to room temperature. The resulting plasma was analysed on board a Beckman Coulter AU680 clinical chemistry analyser using reagents and settings recommended by the manufacturer. Clinical biochemistry for cholesterol analysis was conducted on 8, 22, 27 and 33 week blood samples.

Statistical analysis of mouse data

Data collection, summary calculations and descriptive statistics were carried out using Microsoft Excel 2010. Statistical analysis was carried out using software Graphpad Prism v6. Effects of genotype at different time-points on clinical chemistry parameters were determined by two way ANOVA with repeated measures and Bonferroni's correction for multiple comparisons. Unpaired 2-tailed t tests were used for pairwise comparisons as appropriate and variances were equal, other than in the case of figure 4a male data where a Welch's correction was applied for unequal variance. Data sets were tested for normal distribution and equal variance according to a D'Agostino-Pearson omnibus K2 normality test and a Brown-Forsythe test respectively in Graphpad Prism. Consequently, non-parametric tests were applied as necessary and indicated in the figures. In these cases area under the curve was calculated (baselined to t=0 values) and tested with either a 1-way ANOVA Kruskal-Wallis test and Dunns multiple comparison test where there were multiple groups ($Klf14^{tm1(KOMP)Vlcg}$ knockout mice) or Mann-Whitney 2-tailed t-test when there were only two groups (CRISPR-Cas KO). Further as indicated, individual pairwise comparisons were made using Mann-Whitney 2-tailed t-tests and shown in the figures.

Gene expression in mouse models

Mouse RNAseq data from was collected as described in the Supplemental Note and analysed by the in house bioinformatics team at MRC Harwell using a previously described pipeline[55]. Functional enrichment of mouse differentially expressed genes was assessed using ToppFun[24], with enrichments considered significant after FDR multiple testing correction ($q<0.05$). MEME_chip[55] was used to identify motifs that were significantly over-represented in DNase1 hypersensitivity hotspots upstream of mouse differentiated genes.

Isolation, culture, and differentiation of pre-adipocytes

Abdominal subcutaneous adipose tissue (ASAT) biopsies were obtained from 8 healthy subjects (four females and four males) recruited from the Oxford

BioBank (OBB) (http://www.oxfordbiobank.org.uk), aged 30-50 years with BMI ranging from 22-27kg/m^2. The study was approved by Oxfordshire Clinical Research Ethics Committee (08/H0606/107+5) and all subjects gave written informed consent. Primary pre-adipocytes were isolated as previously described (Collins et al[56]) and cultured in DMEM/F12 Ham nutrient mixture (DMEM/F12), 10% FBS, 2 mmol/L glutamine, 0.25 ng/mL fibroblast growth factor, 100 units/mL penicillin, and 0.1 mg/mL streptomycin. Confluent pre-adipocytes were then stimulated for 14 days with an adipogenic cocktail comprising DMEM/F12, 2mmol/L glutamine, 17mmol/L pantothenate, 100nmol/L human insulin, 10nmol/L triiodo-L-thyronine, 33mmol/L biotin, 10mg/mL transferrin, 1mmol/L dexamethasone, 100 units/mL penicillin, and 0.1mg/mL streptomycin. For the first 4 days, 0.25mmol/L 3-isobutyl-1-methylxanthine and 4mmol/L troglitazone were added. The cells were harvested every day during proliferation and 14 days during differentiation.

RNA isolation and Real-time PCR in pre-adipocytes

Total RNA was extracted from pre-adipocytes as described by Collins et al[56]. Real time PCR was performed on an Applied Biosystems 7900HT, using TaqMan Assays-on-Demand (Applied Biosystems) and Klear Kall Master Mix (KBiosciences). mRNA expression values for target genes were calculated using the DCt transformation method[57]. The DCt was calculated as follows: DCt = efficiency (minimum Ct–sample Ct). Values were normalized to endogenous control genes (PPIA and UBC)[58].

Explant lipogenesis experiments

Adipocytes isolated from abdominal subcutaneous adipose tissue were used to measure activation of insulin-stimulated lipogenesis, as described in detail elsewhere[59]. In brief, isolated human adipocytes were incubated at a concentration of 2% (v/v) in Krebs–Ringer phosphate buffer (pH 7.4) containing albumin (40mg/ml), [3-3H] glucose (5×10^5dpm/ml), unlabelled glucose (1µmol/l) and varying concentrations of human insulin (0-70nmol/l). The incubations were conducted for 2 h at 37°C with air as the gas phase. Incubations were stopped by rapidly chilling the incubation vials to 4°C and the incorporation of radiolabelled glucose into adipocyte lipids (i.e. lipogenesis) was determined. Lipogenesis was expressed as the amount of glucose incorporated either per lipid weight of fat cells or per fat cell number, as described previously[59].

Short Hairpin RNA–Mediated Silencing of KLF14

KLF14 was silenced in primary pre-adipocytes derived from female ASAT. Lentiviral particles were produced by transient co-transfection of HEK293 cells, using KLF14-short hairpin RNA (shRNA) lentiviral transduction construct (SHCLNV-NM_138693.1-255s1c1; Sigma-Aldrich) and non-target shRNA lentiviral constructs (SHC002V; Sigma-Aldrich) with packaging vectors (MISSION [Sigma-Aldrich]). Cells were stably transfected by transduction of pre-adipocytes with lentiviral particles followed by selection in growth media containing 2mg/ml puromycin.

Quantification of intracellular lipid content

Control and KLF14 shRNA cell lysates were prepared in lysis buffer containing 1% IGEPAL-630, 150mM NaCl and 50mM Tris HCl. Lysates were sonicated and an aliquot was collected for protein quantification using the BioRad Dc Protein Assay kit. The remaining lysate was heated at 95°C for 30 minutes, allowed to cool to room temperature, and then centrifuged at 12,000xg for 10 minutes. Triacylglycerol concentration in the supernatant was determined using a commercially available enzymatic kit (Instrumentation Laboratory UK) on an iLAB 650 Chemistry Analyser. Total triacylglycerol was normalised to protein concentration.

Estimation of Cell Doubling Time

Equal number of control and KLF14 shRNA cells were seeded in T75 flasks, $1.5x10^5$ cells. Cells were trypsinized and double counted every 5 days. Doubling time was calculated using the formula $T_d = (t_2 - t_1) \times [\log(2) \div \log(q_2 \div q_1)]$, where t = time (days) and q = cell number.

Insulin stimulated glucose uptake

Glucose uptake was assayed according to the established protocol from a commercial glucose uptake kit (J1342; Promega). The luminescent glucose uptake assay was applied to control and KLF14 shRNA differentiated adipocytes in 96-well plates. In brief, differentiated cells were starved in hormones-free DMEM/F12 medium overnight. The medium was removed and the cells were washed with 100µl of phosphate-buffered saline (PBS) followed with 10nM insulin incubation for 1h at 37°C in 5%CO_2. To initiate glucose uptake, 50µl of 2DG (1mM) in PBS was added to the cells. The uptake reaction was stopped, and the samples were processed as described in the protocol. All assay steps were performed at room temperature. All data were acquired on a PerkinElmer EnSpire 2300 multimode plate reader instrument, with an integration time of 0.5s.

Adipocyte Cell Size and Number Study Population and Sample Collection

Nine females and 9 males carrying KLF14 risk allele and paired control subjects age (30-50) and BMI (22-27kgm^{-2}) matched were recruited from the OBB. Paired abdominal and gluteal subcutaneous adipose specimens were obtained by gun and needle biopsies. The gun biopsies were fixed in 10% paraformaldehyde, embedded in paraffin wax, cut into 5 µm sections, and stained with hematoxylin and eosin. Sections were viewed at 20x magnification, and adipocyte cross-sectional area was calculated using Adobe Photoshop 5.0.1 (Adobe Systems, San Jose, CA) and Image Processing Tool Kit (Reindeer Games, Gainesville, FL). As previously described[60], in order to accurately determine the minimum number of cells required for measurement of cell size distribution in a sample, we took 4 samples and counted 1,000 cells in each. Data were removed by 100 at a time and it was observed that the coefficient of variation started to increase when fewer than 100 cells were included in each biopsy. Therefore, we included only biopsies with more than 100 cells available for quantification (N=9 pairs from each genotype). Statistical significance was assessed using a Wilcoxon signed-rank test. Replication of the cell size was undertaken in a Swedish cohort[61]. Abdominal

subcutaneous adipose specimens were obtained by needle biopsy. Adipocytes were separated from stroma cells by treatment in a shaking bath at 37°C for 60 min with collagenase (0.5mg/L) in 5ml Krebs Ringer phosphate buffer (pH7.4) with purified BSA (40g/L) as previously described[59]. Adipocyte suspensions were then rinsed three times in collagenase-free buffer using nylon filters, and the cell sizes were measured by direct microscopy. The mean adipocyte diameter was calculated from measurements of 100 cells and statistical significance assessed using a Wilcoxon signed-rank test. Cell numbers were calculated using the formula described by Hirsch, et al[62] where DXA measured adipose depot size was divided by mean cell volume.

Date Availability Statement

TwinsUK RNAseq data is available from EGA (Accession number: EGAS00001000805). TwinsUK adipose methylation data is available from ArrayExpress (E-MTAB-1866), and blood methylation data from GEO (GSE50660). TwinsUK genotypes are available upon application to the TwinsUK cohort. METSIM adipose array data is available from GEO(GSE70353).

Methods Only References

44 Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-1089, doi:10.1038/ng.2394 (2012).

45 Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358, doi:10.1093/bioinformatics/bts163 (2012).

46 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-9445, doi:10.1073/pnas.1530509100 (2003).

47 qvalue: Q-value estimation for false discovery rate control. v. R package version 1.40.0.

48 Grundberg, E. *et al.* Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *American journal of human genetics* **93**, 876-890, doi:10.1016/j.ajhg.2013.10.004 (2013).

49 Tsaprouni, L. G. *et al.* Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* **9**, 1382-1396, doi:10.4161/15592294.2014.969637 (2014).

50 Pierce, B. L. *et al.* Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. *PLoS Genet* **10**, e1004818, doi:10.1371/journal.pgen.1004818 (2014).

51 Najafabadi, H. S., Albu, M. & Hughes, T. R. Identification of C2H2-ZF binding preferences from ChIP-seq data using RCADE. *Bioinformatics* **31**, 2879-2881, doi:10.1093/bioinformatics/btv284 (2015).

52 Janky, R. *et al.* iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS computational biology* **10**, e1003731, doi:10.1371/journal.pcbi.1003731 (2014).

53 Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: Finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res* **31**, 3666-3668 (2003).

54 Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-1006, doi:10.1093/nar/gkt1229 (2014).

55 Parsons, M. J. *et al.* The Regulatory Factor ZFHX3 Modifies Circadian Function in SCN via an AT Motif-Driven Axis. *Cell* **162**, 607-621, doi:10.1016/j.cell.2015.06.060 (2015).

56 Collins, J. M., Neville, M. J., Hoppa, M. B. & Frayn, K. N. De novo lipogenesis and stearoyl-CoA desaturase are coordinately regulated in the human adipocyte and protect against palmitate-induced cell injury. *The Journal of biological chemistry* **285**, 6044-6052, doi:10.1074/jbc.M109.053280 (2010).

57 Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).

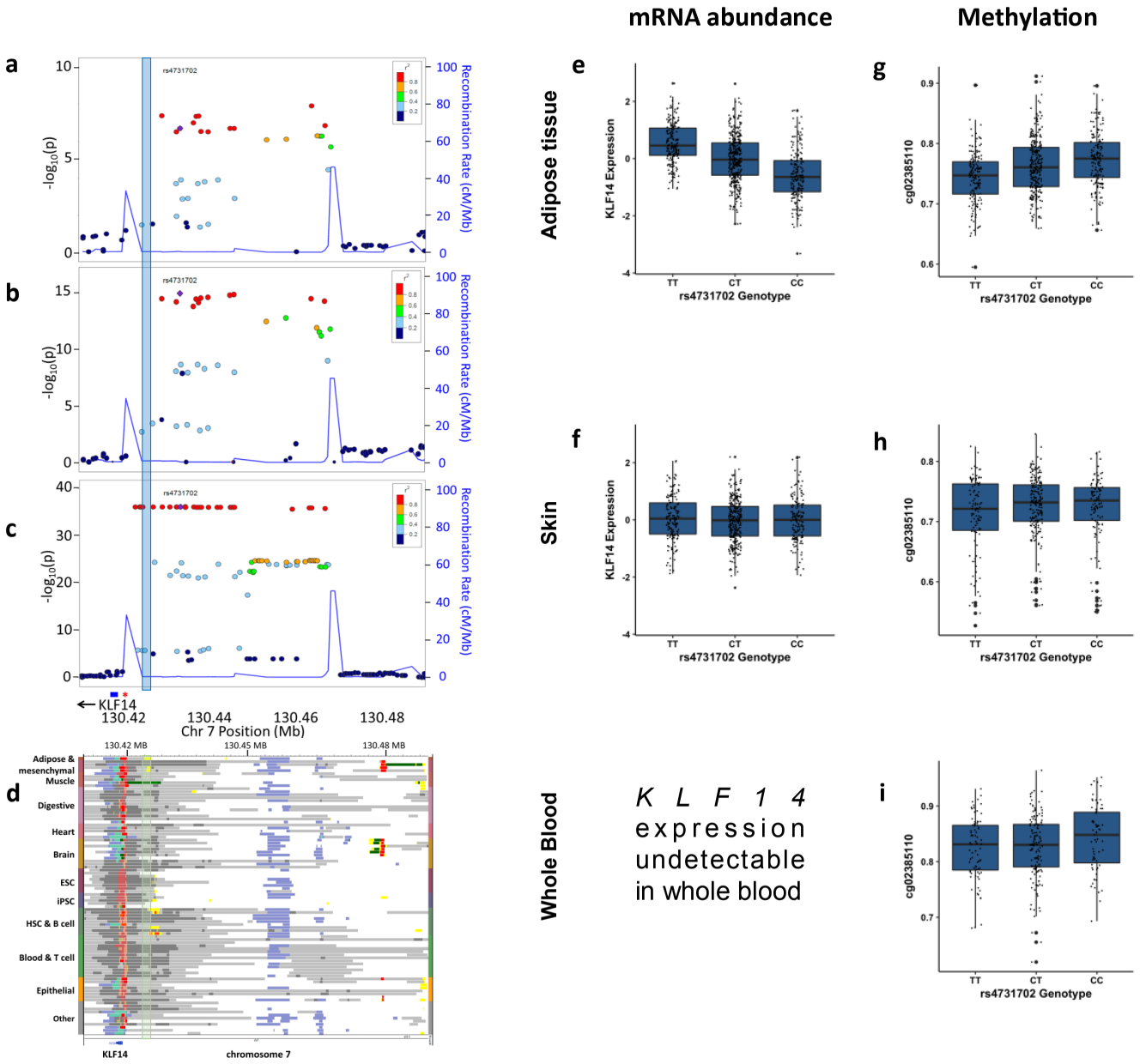
58 Neville, M. J., Collins, J. M., Gloyn, A. L., McCarthy, M. I. & Karpe, F. Comprehensive human adipose tissue mRNA and microRNA endogenous control selection for quantitative real-time-PCR normalization. Obesity **19**, 888-892, doi:10.1038/oby.2010.257 (2011).

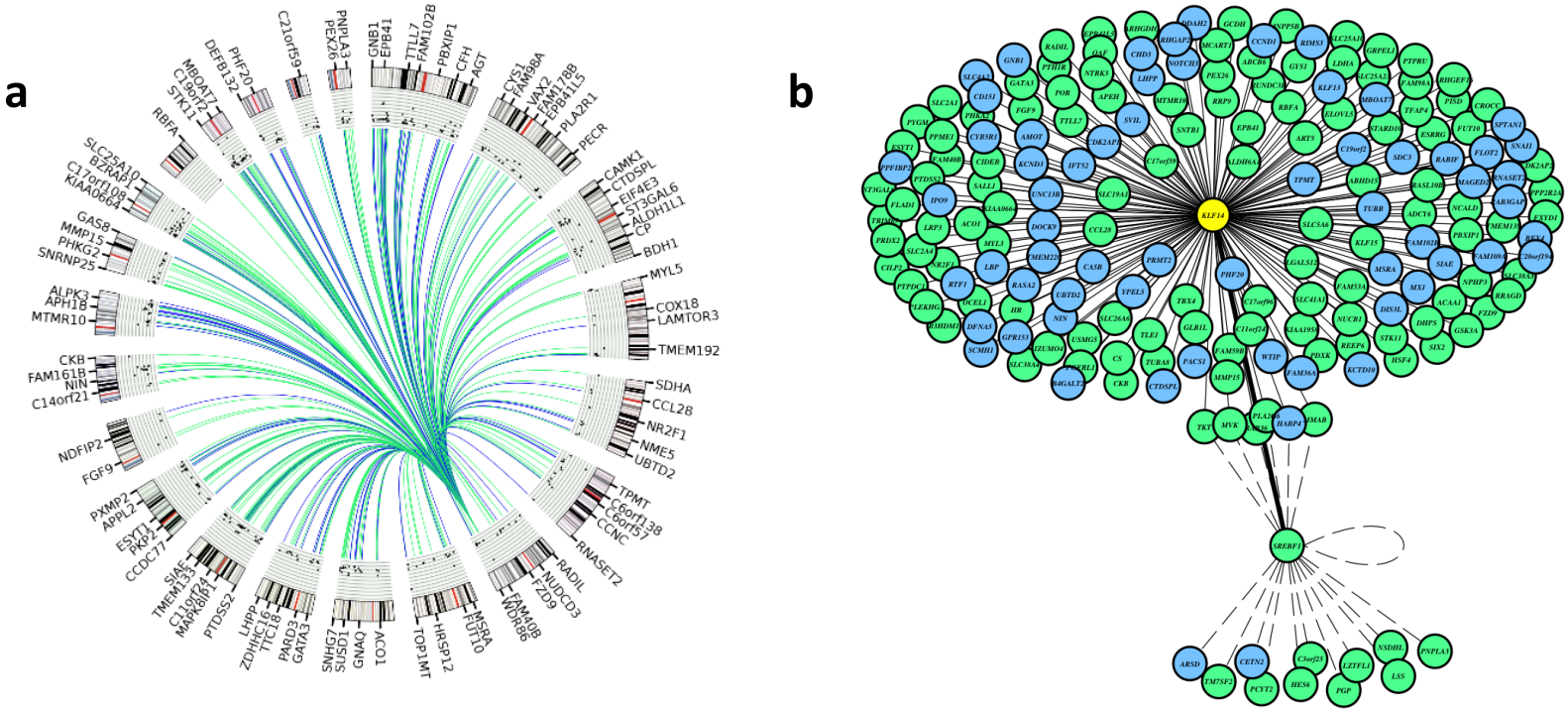
59 Lofgren, P., Hoffstedt, J., Naslund, E., Wiren, M. & Arner, P. Prospective and controlled studies of the actions of insulin and catecholamine in fat cells of obese women following weight reduction. Diabetologia **48**, 2334-2342, doi:10.1007/s00125-005-1961-6 (2005).

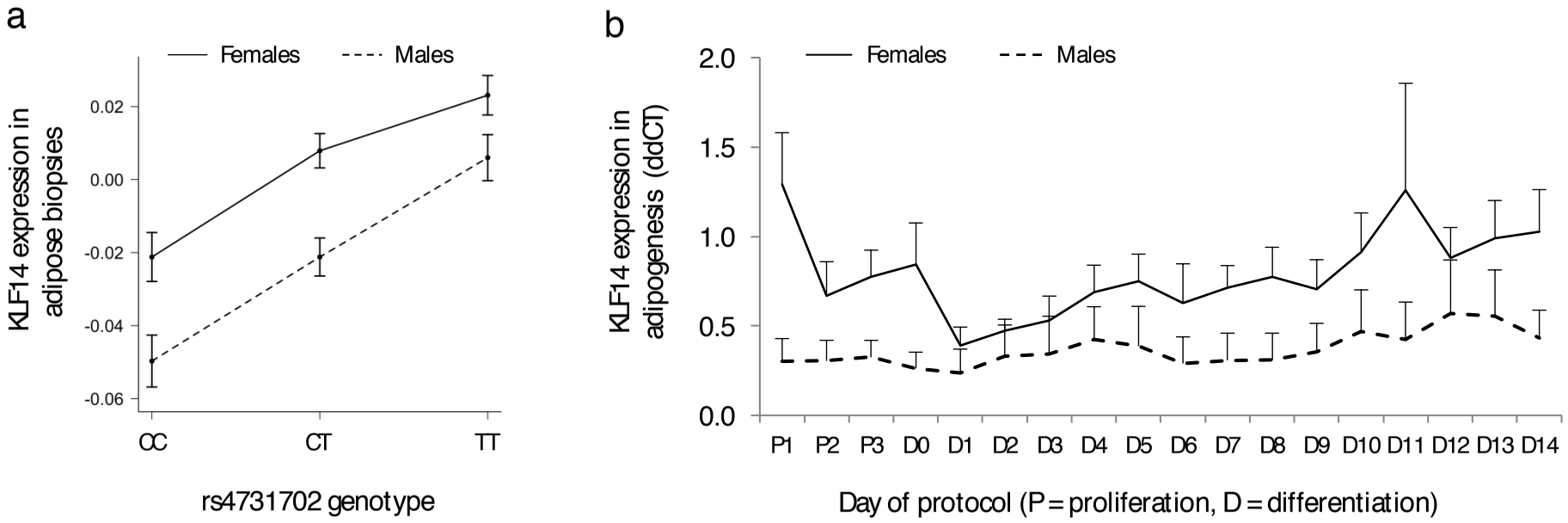
60 Marinou, K. et al. Structural and functional properties of deep abdominal subcutaneous adipose tissue explain its association with insulin resistance and cardiovascular risk in men. Diabetes care **37**, 821-829, doi:10.2337/dc13-1353 (2014).

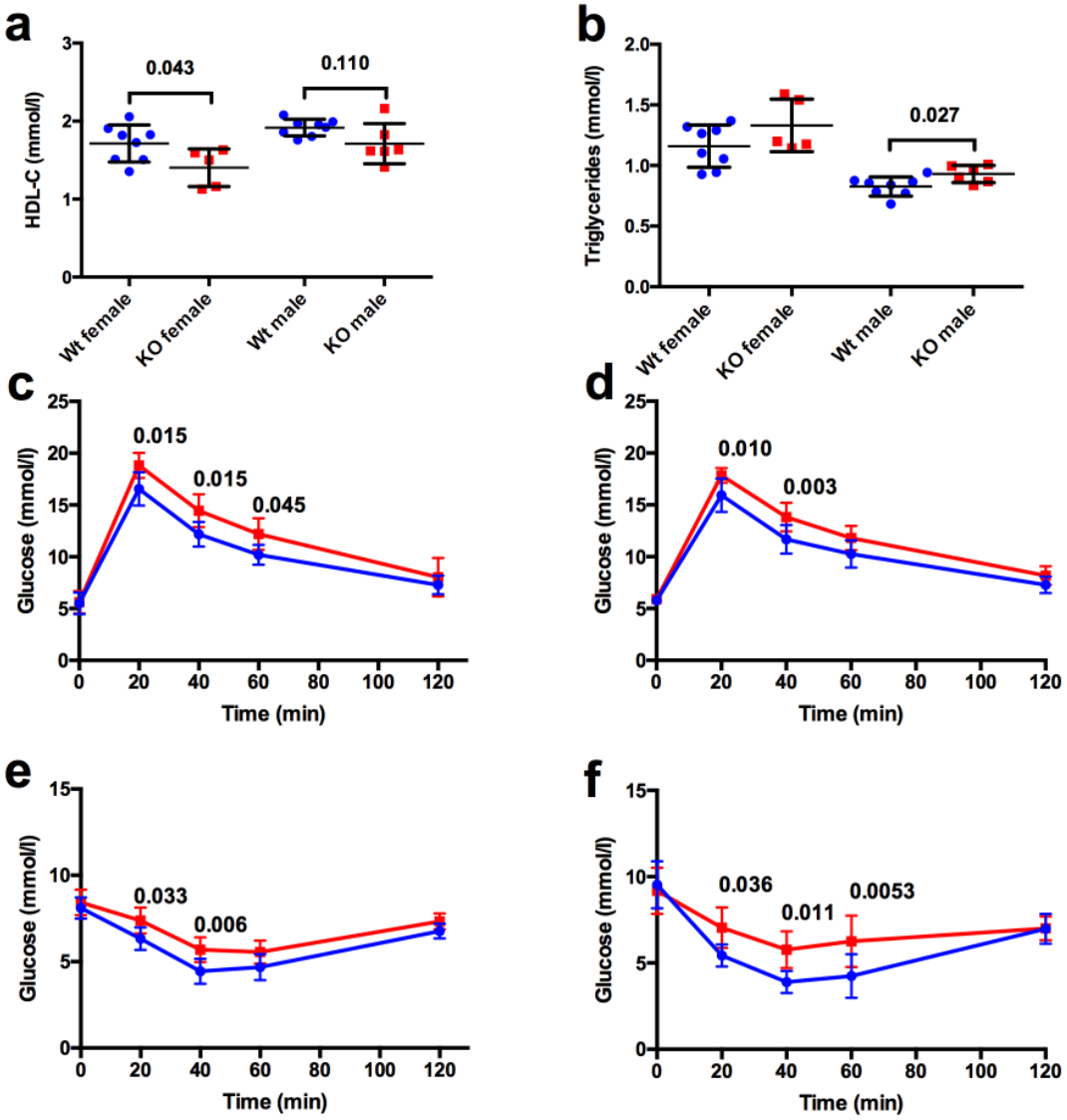
61 Dahlman, I. et al. Numerous Genes in Loci Associated With Body Fat Distribution Are Linked to Adipose Function. Diabetes **65**, 433-437, doi:10.2337/db15-0828 (2016).

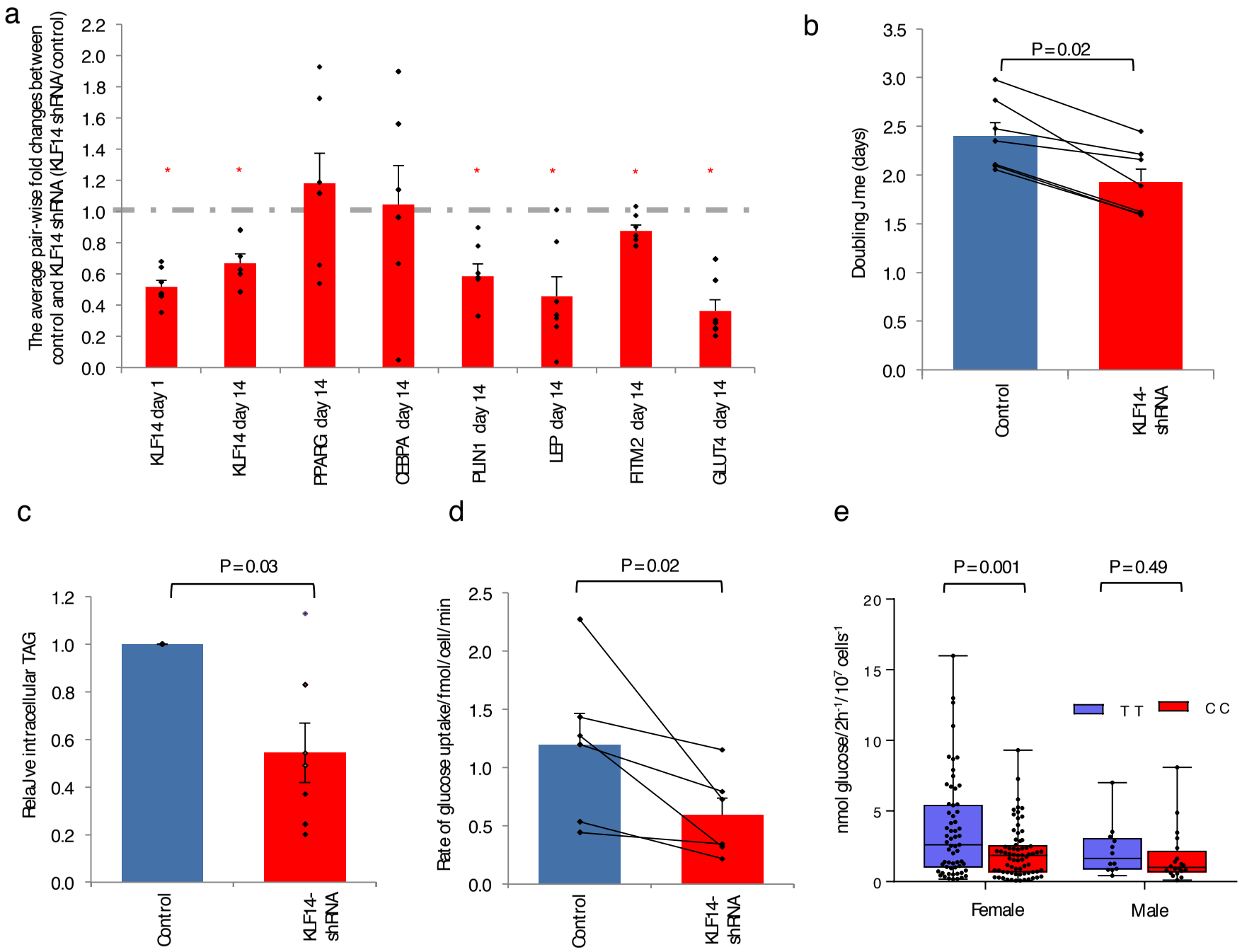
62 Hirsch, J. & Gallian, E. Methods for the determination of adipose cell size in man and animals. Journal of lipid research **9**, 110-119 (1968).

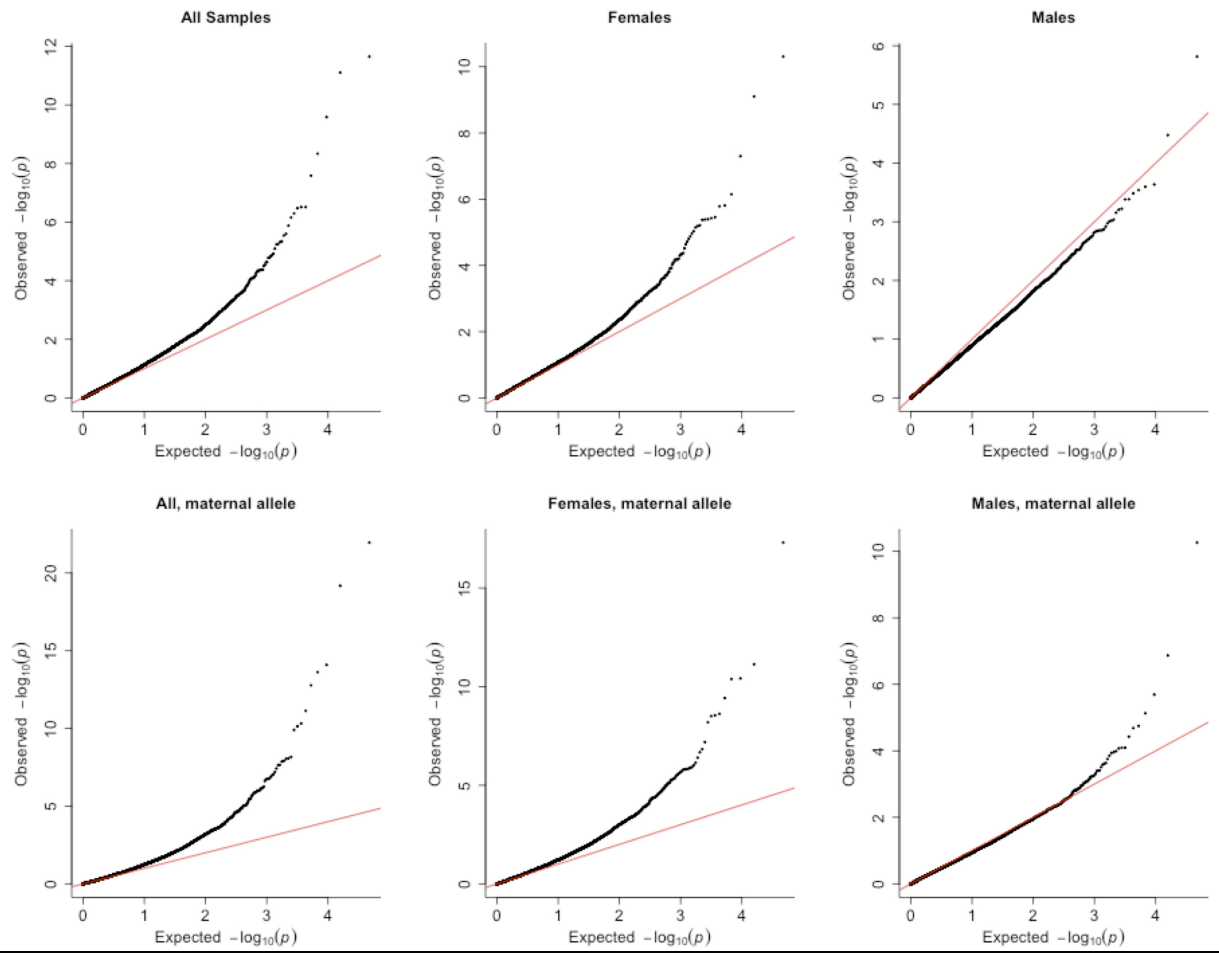

mRNA abundance
Methylation
a
b
c
d
e
f
g
h
i
Adipose tissue
Skin
Whole Blood
-log10(p)
Recombination Rate (cM/Mb)
rs4731702
KLF14
Chr 7 Position (Mb)
130.42
130.44
130.46
130.48
130.42 MB
130.45 MB
130.48 MB
Adipose & mesenchymal
Muscle
Digestive
Heart
Brain
ESC
iPSC
HSC & B cell
Blood & T cell
Epithelial
Other
chromosome 7
KLF14 Expression
cg02385110
TT
CT
CC
rs4731702 Genotype
KLF14 expression undetectable in whole blood


a
b
c
d
METSIM
MGH
DeCODE Maternal
TwinsUK Skin
TwinsUK LCL
TwinsUK Blood
Frequency
P values
KLF14
SREBF1


a
Females
Males
KLF14 expression in adipose biopsies
0.02
0.00
-0.02
-0.04
-0.06
CC
CT
TT
rs4731702 genotype
b
Females
Males
KLF14 expression in adipogenesis (ddCT)
2.0
1.5
1.0
0.5
0.0
P1 P2 P3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 D14
Day of protocol (P = proliferation, D = differentiation)


a
HDL-C (mmol/l)
0.043
0.110
Wt female
KO female
Wt male
KO male
b
Triglycerides (mmol/l)
0.027
Wt female
KO female
Wt male
KO male
c
Glucose (mmol/l)
0.015
0.015
0.045
Time (min)
d
Glucose (mmol/l)
0.010
0.003
Time (min)
e
Glucose (mmol/l)
0.033
0.006
Time (min)
f
Glucose (mmol/l)
0.036
0.011
0.0053
Time (min)

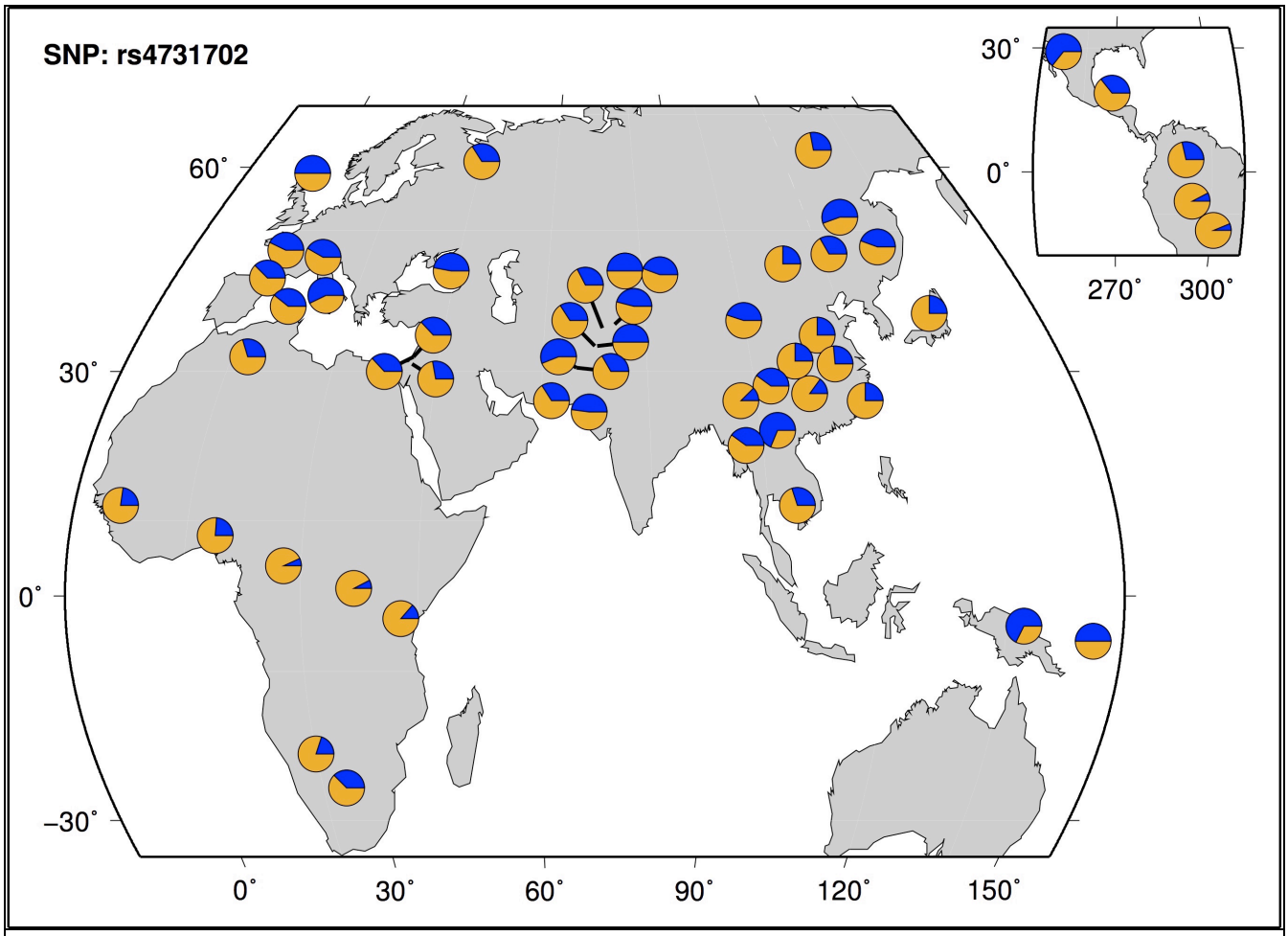

a
The average pair-wise fold changes between control and KLF14 shRNA (KLF14 shRNA/control)
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
*
KLF14 day 1
KLF14 day 14
PPARG day 14
CEBPA day 14
PLIN1 day 14
LEP day 14
FITM2 day 14
GLUT4 day 14
b
P=0.02
Doubling Jme (days)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Control
KLF14-shRNA
c
P=0.03
RelaJve intracellular TAG
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Control
KLF14-shRNA
d
P=0.02
Rate of glucose uptake/fmol/cell/min
2.5
2.0
1.5
1.0
0.5
0.0
Control
KLF14-shRNA
e
P=0.001
P=0.49
nmol glucose/2h-1/10^7 cells-1
20
15
10
5
0
T T
C C
Female
Male



Supplementary Figure 1

Sex-stratified and parent of origin effects on the *KLF14 trans*- eQTL.

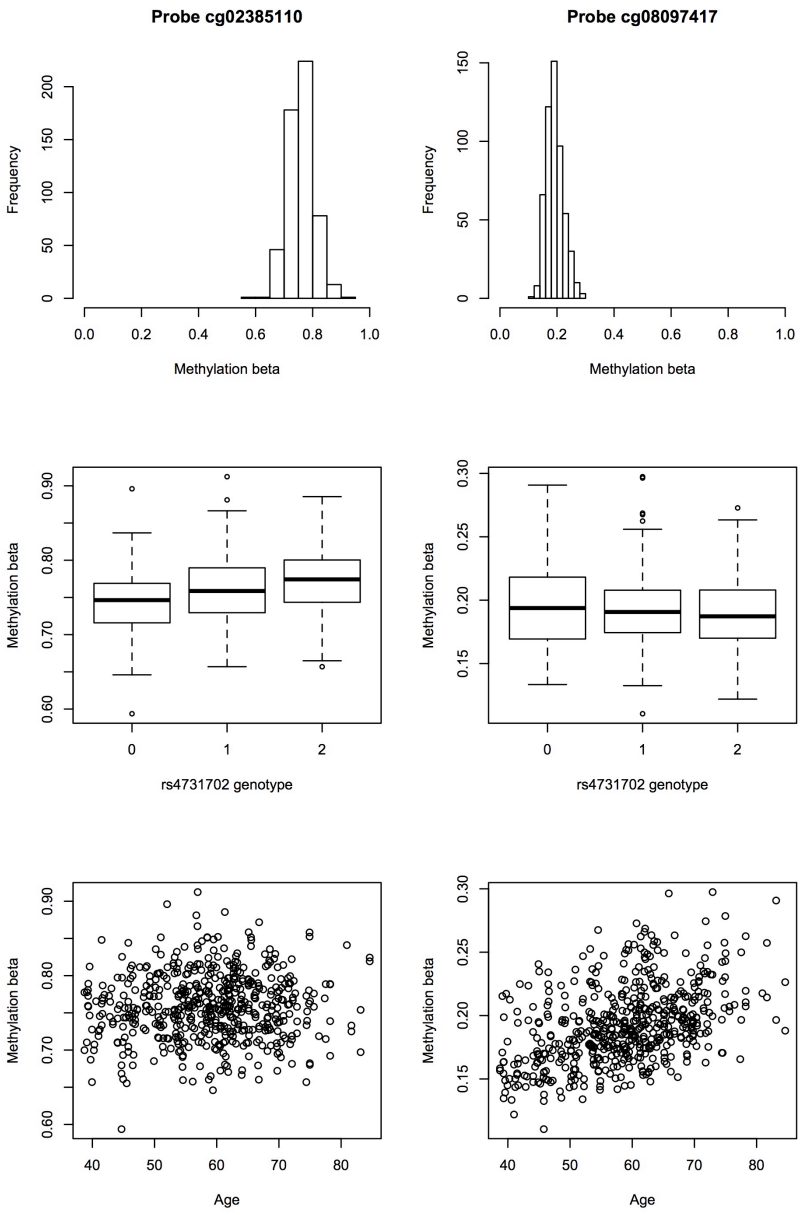
QQ-plots of association between rs473102 and all adipose probes in the deCODE dataset (female N=376 and male N=265). Top row displays results from standard genotypic tests, bottom row displays results for test of association to maternally inherited allele.



Supplementary Figure 2

Geographic distribution of rs4731702 in the Human Genome Diversity Panel.

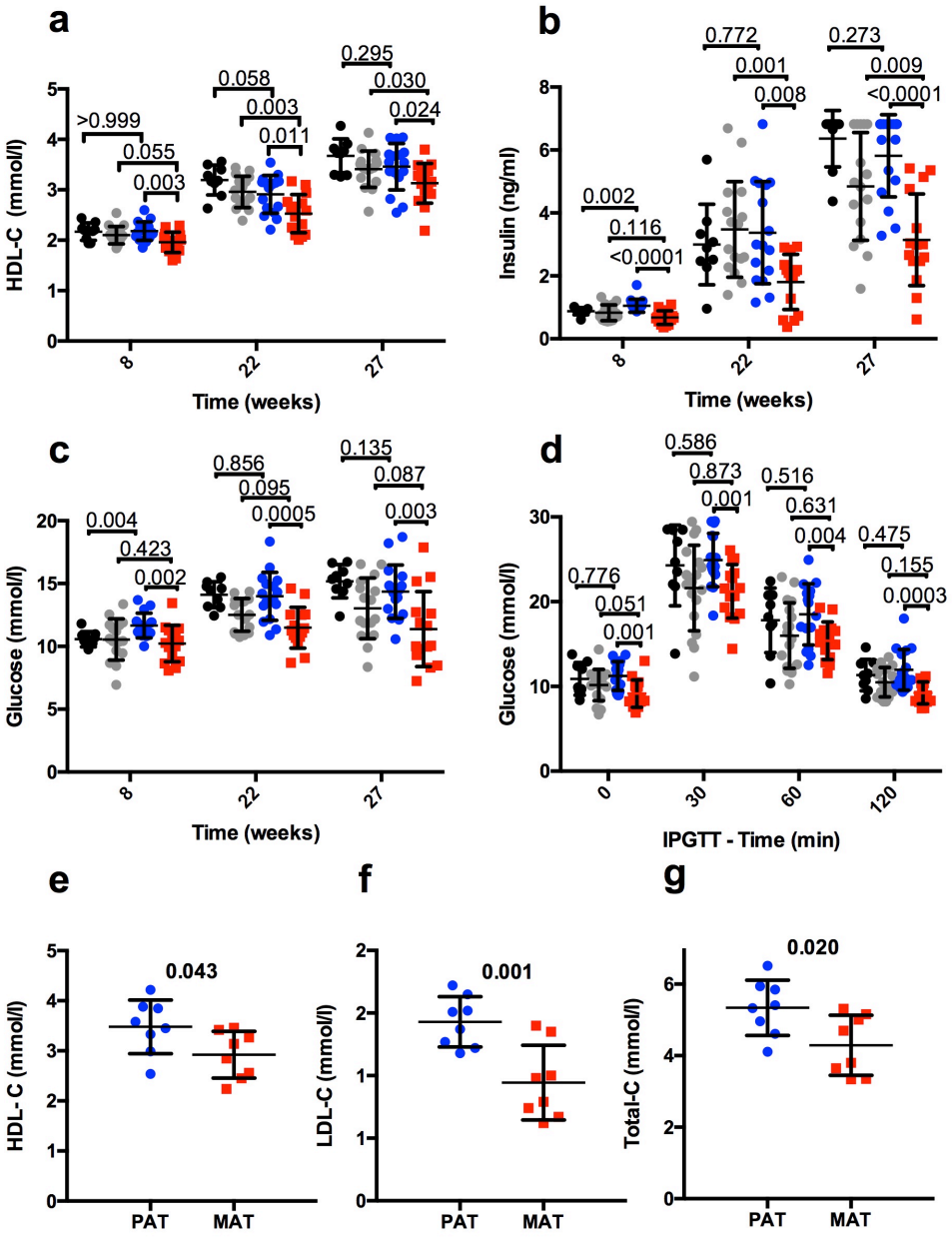
The ancestral T2D risk allele C is colored in orange, the derived allele T in blue. Figure generated from the HGDP Selection Browser at the Pritchard Lab (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/). rs4731702 does not display evidence of positive selection in the HGDP – with global FST in the 80th percentile genome-wide and non-significant iHS and XP-EHH scores in all populations[71]. Tests for recent positive selection at rs4731702 in samples from the United Kingdom using the Singleton Density Score (SDS) were not significant (*P*=0.20), but the trend was towards a recent increase of the non-risk allele T[72]. By contrast, the *KLF14* transcript does exhibit evidence of intolerance to variation in the ExAC exome aggregation dataset[73] (ExAC constraint scores: Missense z= 2.32; Synonymous z = 2.41) suggesting coding changes in *KLF14* are under purifying selection.



Supplementary Figure 3

Distinct 450K methylation probes at the *KLF14 locus* are associated to rs4731702 vs age

Figure shows results for two 450K probes assayed in the TwinsUK adipose samples (N=603). cg02385110, which is ~3KB upstream of *KLF14* is shown on the left, cg08097417 which is at the *KLF14* transcription start site is shown on the right. The beta distributions (top row), association to rs4731702 (T2D risk allele homozygotes = 2) (middle row) and association to age (bottom row) differ between the two probes. cg02385110 has a higher mean beta value, is associated to rs4731702 (P=2.1x10^{-7}), and is not associated to age (P=0.64). In contrast, cg08097417 is not associated to rs4731702 (P=0.99) but is highly associated to age (P=3.6x10^{-61}). Associations were tested with linear mixed effect models adjusting for batch effects, BMI and famlly structure.



Red heterozygous maternally (MAT) inherited KO allele - equivalent to risk alleles

Blue heterozygous Paternally (PAT) inherited KO allele - equivalent to non-risk

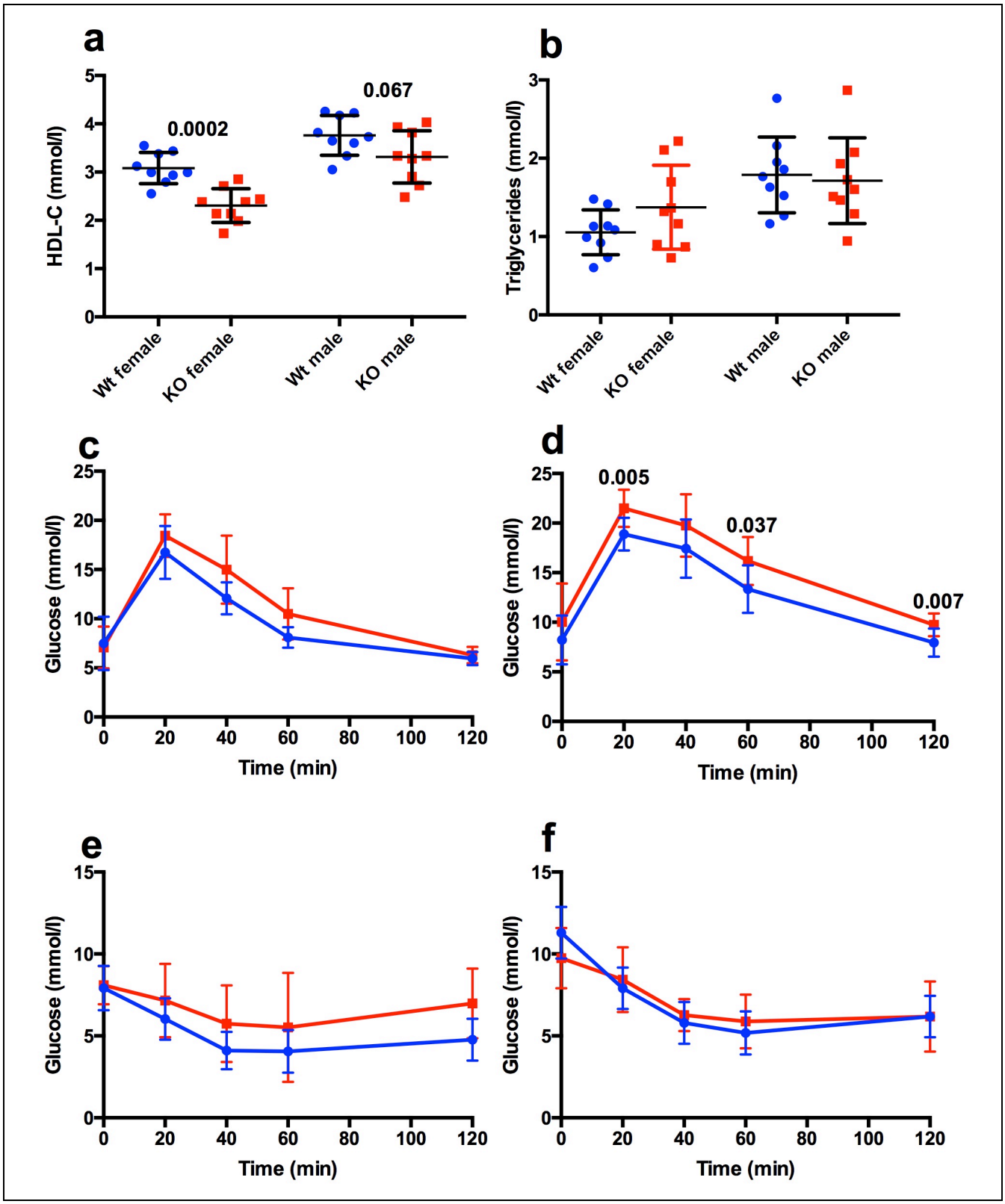
Grey and Black homozygous WILDTYPE (equivalent to non-risk) colonymate mice from the two stocks above:

Grey one wildtype allele from mothers carrying the KO and the other from a wildtype C57BL/6N male
Black one wildtype allele from fathers carrying the KO and the other from a wildtype C57BL/6N female

Supplementary Figure 4

Clinical chemistry analysis of heterozygous *Klf14*$^{tm1(KOMP)Vlcg}$ knockout mice

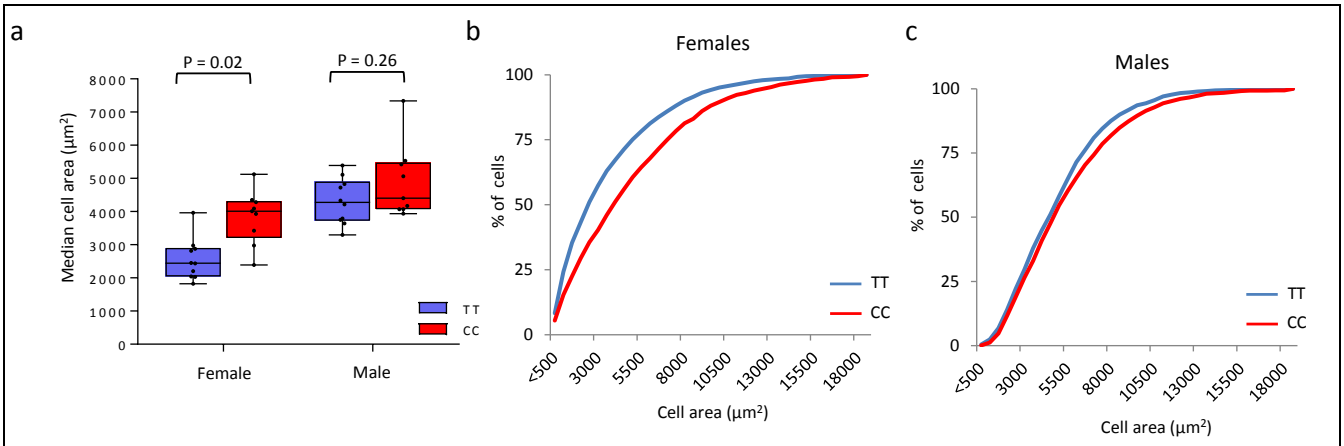
Clinical chemistry parameters were measured in male knockout C57BL/6N *Klf14* mice and their wildtype controls. Since *Klf14* is mono-allelically maternally expressed in mouse and human, we compared heterozygous mice that had inherited the deletion allele from their mother (**heterozygous-MAT**, expressing the deletion) with heterozygous mice that had inherited the deletion allele from their father (**heterozygous-PAT**, not expressing the deletion). We also compared the two groups to their own homozygous wildtype colonymate controls (**wildtype-MAT** and **wildtype-PAT**). This was because we used two separate crosses to produce the MAT and PAT carrier cohorts; one stock from heterozygous mothers and one from heterozygous fathers, each crossed to wildytpe C57BL/6N mice. Mice were fed a standard diet and then switched at 18-weeks of age to a 45kcal% high fat diet. Comparing across the timecourses for **(a)**, HDL-C; **(b)**, insulin; **(c)**, glucose; **(d),** IPGTT by calculating area under the curve, base-lined to t=0 values (data not shown), and analysing with a 1-way ANOVA non-parametric Kruskal-Wallis test and Dunns multiple comparison test we found: that HDL-C **(a)** was lower in MAT compared to either PAT or wildtype-MAT (p= 0.0055 and 0.0086 respectively) and that wildtype-PAT compared to PAT was not significantly (p=0.56) different; that insulin **(b)** was lower in MAT compared to either PAT or wildtype-MAT (p= <0.0003 and 0.0020 respectively) and that wildtype-PAT compared to PAT was not significantly (p>0.99) different; that Glucose **(c)** was lower in MAT compared to PAT (p= <0.0001) and that wildtype–MAT and wildtype-PAT compared to MAT and PAT respectively were not significantly (p=0.79 and >0.99 respectively) different; and that for an IPGTT **(d)** *none* of the group comparisons were significantly different (p>0.99) to each other. Then to examine effects at specific times for **(a)** HDL-C, **(b)** insulin, **(c)** glucose and **(d)** IPGTT individual pairwise comparisons were made at each timepoint using a Mann-Whitney 2-tailed t-test and were significantly reduced in the MAT group compared to PAT group at 8, 22 and 27 weeks in **(a,b,c)** and at all timepoints in **(d)**. The MAT groups were also significantly lower compared to wildtype-MAT for HDL-C and insulin at 22 and 27 weeks. Values are expressed as mean ± SD and in **(a,b,c)** wildtype-MAT n=16, wildtype-PAT n=9, heterozygous-MAT n=15, heterozygous-PAT n=16 and in **(d)** wildtype-MAT n =19, wildtype-PAT n=10, MAT n=16, PAT n=16. For **(e)**, HDL-C; **(f)**, LDL-C and **(g)**, total cholesterol was measured in a 33-week blood sample collected under terminal anaesthetic and was significantly, using an unpaired 2-tailed t-test, reduced in heterozygous-MAT mice compared to PAT mice. Values are expressed as mean ± SD (PAT N =8, MAT N =8). Wildtype MAT grey, wildtype PAT black, MAT KO red, PAT KO blue lines and fill.


a
HDL-C (mmol/l)
0.0002
0.067
Wt female
KO female
Wt male
KO male
b
Triglycerides (mmol/l)
Wt female
KO female
Wt male
KO male
c
Glucose (mmol/l)
Time (min)
d
Glucose (mmol/l)
0.005
0.037
0.007
Time (min)
e
Glucose (mmol/l)
Time (min)
f
Glucose (mmol/l)
Time (min)

Supplementary Figure 5

Clinical chemistry and histological analyses of global CRISPR-Cas9 KO mice

Clinical chemistry parameters were measured in female and male CRISPR-Cas9 knockout (KO) C57BL/6J *Klf14* mice and their wildtype (Wt) controls. Mice were fed a standard diet throughout their lifetimes. **(a)**, HDL-C at 12 weeks of age was significantly reduced in female and reduced with borderline significance in male (unpaired two-tailed t-test). **(b),** triglycerides at 12 weeks of age were not significantly different in females or males (unpaired two-tailed t-test). Comparing across the IPGTT timecourses **(c and d)** by calculating area under the curve, base-lined to t=0 values, (data not shown) and analysing using a Mann-Whitney two-tailed t-test showed there was no significant difference between male or female KO and wildtype mice (p=0.23 and 0.93 respectively). Then in order to make comparisons at each timepoint for **(c and d),** individual pairwise Mann-Whitney 2-tailed t-tests were carried out, and as for AUC no differences were found in the female data, although in males glucose levels were nominally higher at 20, 60 and 120 minutes of the test. **(e and f)**, ITT at 16 weeks was not significantly changed between wildtype and KO groups of either female (e) or male (f) mice, analysed by 2-way ANOVA with repeated measures and Bonferroni correction. Values are expressed as mean ± SD and in **(a,b)** female wildtype n=9, female KO n=9, male Wt n=9, male KO n=9 and in **(c,d,ef)** female wildtype n=9, female KO n=8, male Wt n=9, male KO N=9. Wildtype mice in blue lines and fill, KO mice in red lines and fill.



Supplementary Figure 6

Gluteal adipose tissue of T2D risk allele homozygotes contains fewer, larger mature adipocytes compared to non-risk allele homozygotes**.**

Adipocyte cell area in histological sections of subcutaneous gluteal adipose biopsies from the Oxford BioBank. **a,** Median cell area in female (N=18, mean ± SEM) and male (N=18, mean ± SEM) gluteal biopsies stratified by genotype at rs4731702. **b,** Cumulative frequency distribution of adipocyte cell area in females (N=18). **c,** Cumulative frequency distribution of adipocyte cell area in males (N=18). Statistical significance was assessed using a Wilcoxon signed-rank two-sided test.

Supplementary Material

Regulatory variants at *KLF14* influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition

Contents:

Supplementary Note

Natural selection at *KLF14*

Despite the geographic range in allele frequency, rs4731702 does not display evidence of positive selection by several tests. rs4731702 is not under positive selection in the Human Genome Diversity Panel (HGDP) – with global FST in the 80th percentile genome-wide and non-significant iHS and XP-EHH scores in all populations[1]. Tests for recent positive selection at rs4731702 in samples from the United Kingdom using the Singleton Density Score (SDS) were not significant (*P*=0.20), but the trend was towards a recent increase of the non-risk, derived allele T[2]. By contrast, the *KLF14* transcript does exhibit evidence of intolerance to variation in the ExAC exome aggregation dataset[3] (ExAC constraint scores: Missense z= 2.32; Synonymous z = 2.41) suggesting coding changes in *KLF14* are under purifying selection.

TwinsUK Data Extended Methods

Biopsy collection Biopsies and blood samples from 856 healthy female twins from the TwinsUK cohort were collected within the MuTHER project[4]. In short, subcutaneous adipose tissue punch biopsies from a photo-protected area of the stomach adjacent and inferior to the umbilicus were obtained from consented individuals and dissected to yield adipose and skin biopsies. Peripheral blood samples were also collected as part of the study and Lymphoblastoid Cell Lines (LCLs) generated via transformation of the B-lymphocyte fraction with Epstein-Barr Virus (EBV).

Genotypes The TwinsUK samples were genotyped on a combination of platforms (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo Illumina arrays) and samples were imputed using the 1000 Genomes[5] phase 1 reference panel using IMPUTE2[6] as described previously[7].

RNA-sequencing and quantification Adipose, skin, LCL and whole blood TwinsUK RNA samples were sequenced and quantified as previously described[7]. In short, reads were mapped to Gencode version 10 and variation in sequencing depth between samples was corrected by normalizing the number of reads to the median number of well-mapped reads. We used only exons that were quantified in more than 90% of the individuals. Exons were rank-normal transformed for downstream analysis.

Preparation, quality control and genetic analysis of large cohorts used in Sex x SNP interaction analysis.

UK Biobank: The UK Biobank recruited more than 500,000 people aged 37-73 years (99.5% were between 40 and 69 years) from across the country in 2006-10. Participants provided a range of information via questionnaires and interviews (such as demographics, health status, and lifestyle); anthropometric measurements, blood pressure readings, and blood, urine and saliva samples were taken for future analysis. This has been described in more detail elsewhere[8]. Genotype data from the May 2015 release was available for a subset of 152,249 participants from UK Biobank. In addition to the quality control metrics performed centrally by UK Biobank, we defined a subset of "white European" ancestry samples (n=120,286) as those who both self-identified as white British and were confirmed as ancestrally "Caucasian" using principal components analyses of genome-wide genetic information. A maximum of 118,193 individuals (62,165 females and 56,027 males) with genotype and valid BMI, height, waist, hip, and type 2 diabetes measures were available for downstream analyses. Prevalent type 2 diabetes status was defined using self-reported medical history and medication in UK Biobank participants[9]. The genotypes of the genetic variant were extracted from UK Biobank's imputation dataset. We excluded individual genotypes if the genotype probability was less than 0.9. We converted the BMI, height, waist, and hip measures to a normal distribution by inverse normalising the variables. Each analysis was adjusted for covariates: age, age^2, sex, six (within UK) ancestry principal components, and array used to measure genotypes.

GERA: The Resource for Genetic Epidemiology on Adult Health and Aging (GERA) is a large multi-ethnic population-based cohort, created for investigating the genetic and environmental basis of age-related diseases [dbGaP phs000674.p1]. T2D status is based on ICD-9 codes in linked electronic medical health records, and quality control of these data have been previously described[10]. We extracted 6,961 T2D cases and 13,922 controls that were genotyped using a custom array to maximise coverage of common and low-frequency variation in non-Hispanic whites[11]. We used multi-dimensional scaling in PLINK[12] to obtain three principal components to correct for population structure. The autosomal genotype scaffold was then pre-phased using SHAPEITv2.5[13]. The resulting haplotypes were imputed, using an updated implementation of IMPUTE[6], up to the Haplotype Reference Consortium panel[14]. The interaction of sex with rs4731702 was evaluated in a logistic regression model, with adjustment for the three principal components as covariates.

WTCCC/ UKT2D (The Wellcome Trust Case Control Consortium/United Kingdom Type 2 Diabetes Genetics consortium). Details of the samples and quality control of GWAS data have been previously described[15],[15]. The clean genotype data was pre-phased using SHAPEITv2.5[13]. The resulting haplotypes were imputed, using an updated implementation of IMPUTE[6], up to the Haplotype Reference Consortium panel[14]. The genotypes of the genetic variant were extracted from imputed dataset. We excluded individual genotypes if the genotype probability was less than 0.9. The interaction of sex with rs4731702 was evaluated in a logistic regression model, with adjustment for sex as covariate.

Construction of Mouse Models

Klf14$^{tm1(KOMP)Vlcg}$ **Deletion Mice** The mouse strain (*Klf14*$^{tm1(KOMP)Vlcg}$) used in this project was generated from targeted embryonic stem (ES) cells for *Klf14* obtained from the KOMP repository www.komp.org, a NCRR-NIH supported mouse repository. The U42-RR024244 ES cells from which this mouse was generated were created by Velocigene from funds provided by the trans-HIH Knock-out Mouse Project (KOMP) (Grant number 5U0101HG004085). Live mice were imported on a C57BL/6NTac USA background and rederived into the MRC Harwell Mary Lyon Centre specified pathogen free (SPF) facility and maintained on C57BL/6NTac. *Klf14*$^{tm1(KOMP)Vlcg}$ mice were kept and studied in accordance with UK Home Office legislation and local ethical guidelines issued by the Medical Research Council (Responsibility in the Use of Animals for Medical Research, July 1993; home office license 30/3146).

CRISPR-Cas9 whole-body knockout The CRISPR-Cas9 whole-body knockout mouse was generated using a guide RNA (gRNA) designed to target the beginning of the *Klf14* gene (**Supplementary Table 10**). In vitro transcribed Cas9 mRNA (100 ng/µl; TriLink BioTechnologies) and gRNA (50 ng/µl) were injected into cytoplasm of fertilized oocytes from C57BL/6J mice. Genomic DNA samples from founder mice were screened for frameshift mutations by PCR and confirmed by Sanger sequencing (primers in **Supplementary Table 11**). The most commonly detected mutant allele was a 7-bp deletion, 5'-GAGTGCC-3', and founder mice with this allele were selected for breeding to obtain litters with both homozygous knockout mice and wild-type mice for experiments.

CRISPR-Cas9 conditional knockout The CRISPR-Cas9 conditional knockout mouse was generated using two gRNAs designed to target sites upstream and downstream of the *Klf14* gene; two single-strand DNA oligonucleotides bearing loxP sequences along with 80-nt homology arms matching the target sites were synthesized (IDT) (**Supplementary Table 10**). In vitro transcribed Cas9 mRNA (100 ng/µl) and gRNAs (25 ng/µl each) and single-strand DNA oligonucleotides (100 ng/µl each) were injected into cytoplasm of fertilized oocytes from C57BL/6J mice. Genomic DNA samples from founders were screened for correct loxP sequences flanking the *Klf14* gene by PCR and confirmed by Sanger sequencing (primers in **Supplementary Table 11**). Offspring of founder mice with correct loxP sequences were screened to identify those in which the two loxP sequences segregated on the same chromosome. Those mice were bred for multiple generations with *Adipoq*-Cre mice of the C57BL/6J background [B6;FVB-Tg(Adipoq-cre)1Evdr/J, The Jackson Laboratory] to obtain litters with both homozygous *Klf14* loxP knock-in mice and wild-type mice that were also positive for *Adipoq*-Cre for experiments. The presence of the *Adipoq*-Cre allele was confirmed by PCR (primers in **Supplementary Table 11**).

Gene Expression in Mouse Models

Klf14$^{tm1(KOMP)Vlcg}$ **Deletion Mice** RNA was extracted from snap frozen subcutaneous tissue, isolated at the same time of day, using phenol chloroform and a RNeasy Mini Kit (Qiagen, UK) according to the manufacture's instructions. RNA integrity was assessed using a 2100 Bioanalyzer (Agilent Technologies) and samples with a RNA

integrity (RIN) number of 8 and above (N=4/group) were selected for RNASeq analysis. These samples were sent to the Oxford Genomics centre where a quality control report was generated and a library of read data compiled. Differentially expressed genes were identified by comparing the RNAseq profiles of male mice inheriting the *Klf14*$^{tm1(KOMP)Vlcg}$ deletion allele from their mother (resulting in no *Klf14* expression) to male mice inheriting the deletion allele from their father (maintained expression from the maternally inherited non-deletion allele).

CRISPR-Cas9 conditional knockout RNA was extracted from snap-frozen subcutaneous adipose tissue from 4-month-old littermates (n = 2 females, 2 males each of wild-type and adipose-specific knockout mice) using TRIzol RNA Isolation Reagent according to the manufacturer's instructions (Thermo Fisher Scientific). Library generation and RNA sequencing were performed at the Bauer Core Facility of Harvard University. On average, about 25 million 75-bp single-end reads were obtained for each sample on the Illumina NextSeq 500 platform. Differentially expressed genes were identified by comparing the RNAseq profiles of knockout mice to wile-type mice.

Supplementary Tables

Cohort	Samples	Platform	Female:Male	Population	Citation
TwinsUK	776	RNAseq	100:0	Healthy female twins from the United Kingdom	Buil *et al,* Nature Genetics 2015
METSIM	770	Affymetrix U219 array	0:100	Healthy men from Finland	Civelek *et al*, AJHG 2017
deCODE	589	Agilent array	58:42	Healthy men and women from Iceland	Emilsson *et al*, Nature, 2008
MGH	701	Agilent array	75:25	Obese bariatric surgery patients from USA	Greenawalt *et al*, Genome Research 2011

Supplementary Table 2| Characteristics of subcutaneous adipose gene expression cohorts used in *trans*-network replication

Tissue	Source	Platform	N	*Cis-eQTL P*	*Trans-EqtI present*
Subcutaneous Adipose	TwinsUK	RNAseq	776	1.8×10^{-36}	Yes
Subcutaneous Adipose	METSIM	Microarray	770	2.4×10^{-6}	Yes
Subcutaneous Adipose	deCODE	Microarray	589	2.6×10^{-10}	Yes
Subcutaneous Adipose	deCODE maternal	Microarray	589	6.6×10^{-20}	Yes
Subcutaneous Adipose	MGH, Greenawalt et al	Microarray	701	1.6×10^{-11}	Yes
Omental Adipose	MGH, Greenawalt et al	Microarray	848	3.6×10^{-7}	Yes
Subcutaneous Adipose	GTEx	RNAseq	298	0.0001	Yes
Visceral Adipose	GTEx	RNAseq	185	0.97	Yes
Skin	TwinsUK	RNAseq	672	0.79	No
Lymphoblastoid Cell Line	TwinsUK	RNAseq	765	Not expressed	No
Whole Blood	TwinsUK	RNAseq	368	Not expressed	No
Skeletal Muscle	Keildson et al	Microarray	200	Not significant	No
Skeletal Muscle	GTEx	RNAseq	361	Not significant	No
Pancreatic Islets	Van de Bunt et al	RNAseq	118	Not significant	No
Monocytes	Zeller et al	Microarray	1,490	Not significant	NA
Liver	MGH, Greenawalt et al	Microarray	651	Not significant	No
Liver	Innocenti et al	Microarray	266	Not significant	NA
Whole Blood	Westra et al	RNAseq & Microarray	5,311	Not significant	No
Monocytes/Macrophages	Cardiogenics	Microarray	758	Not significant	NA
Dendritic Cells	PhenoGenetic	Microarray	534	Not significant	NA
CD4+ T Cells	Raj et al	Microarray	407	Not significant	NA
CD19+ B cells	Fairfax et al	Microarray	283	Not significant	NA
Lung	Hao et al	Microarray	1,111	Not significant	NA
10 Brain regions	UK Brain Expression Consortium, Ramasamy et al	Microarray	134	Not significant	No
Remaining 41 GTEx tissues	GTEx	RNAseq	70-338	Not significant	No

Supplementary Table 3 | *KLF14 cis*-eQTL is limited to adipose tissue. Association between *KLF14* expression and T2D associated SNP rs4731702. Most published studies only report summary statistics for significant eQTLs – 'Not significant' means the study did not find a *cis*-eQTL between rs4731702 *KLF14,* in some cases this may mean *KLF14* was not expressed in that tissue. The *trans*-eQTL was considered present if at least one *trans*-gene had $P < 0.05$, Bonferonni corrected for the number of *trans*-genes quantified in that dataset. Datasets where *trans*-analysis or raw genotypes were unavailable are marked 'NA'.

Test	Sample	*P*	β	Standard Error
Standard	All	2.6×10^{-10}	-0.023	0.0036
Maternal Allele	All	6.6×10^{-20}	-0.046	0.0049
Paternal Allele	All	0.89	-0.001	0.0051
Standard	Females	1.1×10^{-05}	-0.020	0.0045
Standard	Males	1.5×10^{-06}	-0.027	0.0056
Maternal Allele	Females	4.1×10^{-11}	-0.044	0.0064
Maternal Allele	Males	5.5×10^{-11}	-0.051	0.0074
Paternal Allele	Females	0.96	-0.0003	0.0066
Paternal Allele	Males	0.85	0.002	0.0077

Supplementary Table 4 | Sex-stratified and parent of origin effects on the *KLF14 cis*-eQTL. Association between rs4731702 and adipose expression of *KLF14* in the deCODE dataset (female N=376 and male N=265). The T2D risk allele C is the reference allele.

	KLF14 expression x Trait		Genes associated to trait (%)		Enrichment for association in *KLF14* network	
Trait	***P***	**Direction**	***KLF14* Network genes**	**All expressed genes excluding *KLF14* Network**	**OR**	***P***
Combined Insulin Resistance Phenotype*	1.08×10^{-3}	-	151 (39%)	5,261 (28%)	1.66	1.82×10^{-6}
Fasting Insulin	8.08×10^{-6}	-	200 (52%)	7,254 (39%)	1.70	1.88×10^{-7}
Fasting Glucose	1.63×10^{-6}	-	10 (3%)	301 (2%)	1.63	0.10
HOMA-IR	5.73×10^{-7}	-	200 (52%)	7,234 (39%)	1.73	8.82×10^{-8}
BMI	0.90	-	373 (97%)	16,472 (88%)	4.38	9.08×10^{-10}
HDL	0.04	+	300 (78%)	12,014 (65%)	1.96	8.28×10^{-9}
LDL	0.76	-	3 (0.78%)	68 (0.36%)	2.16	0.17
Triglycerides	0.51	-	267 (70%)	10,585 (57%)	1.76	1.78×10^{-7}
Whole Body Fat	0.12	-	139 (36%)	4,680 (25%)	1.68	1.47×10^{-6}
Waist-hip ratio	0.09	-	140 (36%)	4,511 (24%)	1.81	5.97×10^{-8}
Android/Gynoid Ratio	1.75×10^{-3}	-	314 (82%)	12,706 (68%)	2.12	9.55×10^{-10}
Adiponectin	0.90	-	46 (12%)	1,295 (7%)	1.81	3.313×10^{-4}
Leptin	0.65	-	13 (4%)	249 (2%)	2.60	2.60×10^{-3}

Supplementary Table 7 | Expression of *KLF14* and the 385 *trans*-genes are enriched for association to concurrently measured metabolic traits. Association between adipose expression in TwinsUK (N = 776 females) and concurrently measured metabolic traits. All associations were adjusted for body-mass index (BMI), except to BMI itself. In *KLF14* expression vs Trait columns, *P* reports the p value of the association between *KLF14* expression and the corresponding trait and Direction reports whether the association was direct (+) or inverse (-). Genes associated to trait columns list the number of genes in each category which were associated to the respective trait at FDR 5%. Significance of enrichment for trait-association in the *KLF14* network genes compared to all adipose genes was assessed using a one-tailed Fisher's exact test. OR, odds ratio; BMI, body-mass index, HDL, high density lipoprotein; LDL, low density lipoprotein. *Association with insulin resistance was assessed via a combined insulin resistance phenotype consisting of association with fasting insulin, HDL and triglycerides with a directionality consistent with risk or protection from a phenotype of insulin resistance.

Guide RNA	Protospacer	PAM	Single-strand olionucleotide sequence
Klf14 whole-body knockout	5’-GGACATAGACACCAGGCACT-3’	5’-CGG-3’	N/A
Klf14 conditional knockout upstream	5’-GCGGTGATCTATTGGCGACA-3’	5’-AGG-3’	5’GCGAAACAAAGTTTTCTCATTTAGCAGTTCGGCTGTTAATGCAGGATCGGC AGCCTTGTCGAATTCATAACTTCGTATAATGTATGCTATACGAAGTTATGCCA ATAGATCACCGCGTAGCCATTAATTAATAGCGTTTTGAATCAGAAACGGTTTT AAC-3’
Klf14 conditional knockout downstream	5’-GTCGGACTCTGGAGAGGGAC-3’	5’-TGG-3’	5’CGTTCATTTCAGAAGATGTTTGTGTTGCAGGACTTCCTGCTATCTTCTGGCT CCCAGTCCGAATTCATAACTTCGTATAATGTATGCTATACGAAGTTATCTCTC CAGAGTCCGACAAGTTTGTCTGAACCCATCCCCAGTGCGGAGCAACCAGTCT GTA-3’

Supplementary Table 10 | Guide RNAs used in CRISPR-Cas9 mouse models

PCR primer	Oligonucleotide sequence
Klf14 whole-body knockout forward	5’-AGCTCGTCTGGCTCCAAG-3’
Klf14 whole-body knockout reverse	5’-GGGCACCACAGCTTAAATCA-3’
Klf14 conditional knockout upstream forward	5’-TTCTATTTCAAGCGGGGATG-3’
Klf14 conditional knockout upstream reverse	5’-CTTGTCTGCGCTTTCTCTCC-3’
Klf14 conditional knockout downstream forward	5’-GCAGGCTGTTTGGAAGAAAC-3’
Klf14 conditional knockout downstream reverse	5’-CATGGTGGAAAGCTCTGGTT-3’
Adipoq-Cre forward	5’-GGATGTGCCATGTGAGTCTG-3’
Adipoq-Cre reverse	5’-ACGGACAGAAGCATTTTCCA-3’

Supplementary Table 11 | PCR primers used in CRISPR-Cas9 mouse models

Supplementary References

1 Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19**, 826-837, doi:10.1101/gr.087577.108 (2009).

2 Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760-764, doi:10.1126/science.aag0776 (2016).

3 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).

4 Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-1089, doi:10.1038/ng.2394 (2012).

5 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

6 Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529, doi:10.1371/journal.pgen.1000529 (2009).

7 Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* **47**, 88-91, doi:10.1038/ng.3162 (2015).

8 Collins, R. What makes UK Biobank special? *Lancet* **379**, 1173-1174, doi:10.1016/S0140-6736(12)60404-8 (2012).

9 Eastwood, S. V. *et al.* Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank. *PloS one* **11**, e0162388, doi:10.1371/journal.pone.0162388 (2016).

10 Cook, J. P. & Morris, A. P. Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility. *European journal of human genetics : EJHG* **24**, 1175-1180, doi:10.1038/ejhg.2016.17 (2016).

11 Hoffmann, T. J. *et al.* Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics* **98**, 79-89, doi:10.1016/j.ygeno.2011.04.005 (2011).

12 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).

13 O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* **10**, e1004234, doi:10.1371/journal.pgen.1004234 (2014).

14 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).

15 Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* **42**, 579-589, doi:10.1038/ng.609 (2010).